Quantitative Linguistic Analysis of Czech Sign Language

Jiří Langer
Jan Andres
Martina Benešová
Dan Faltýnek

Quantitative Linguistic Analysis of Czech Sign Language

Jiří Langer
Jan Andres
Martina Benešová
Dan Faltýnek

Not for sale.

Dedicated to the memory of Gabriel Altmann (1931–2020),
Peter Grzybek (1957–2019) and Luděk Hřebíček (1934–2015),
the pioneers of quantitative linguistics.

CONTENTS

PREFACE

The main aim of this book is to present current research outcomes from quantitative analysis of Czech sign language. A multidisciplinary research project entitled "The Theoretical Basis for Teaching Czech Sign Language Tested through Quantitative Linguistic Methods", funded by the Czech Science Foundation, was carried out by researchers from three faculties of Palacký University in Olomouc (Czech Republic). It assumes that a deeper linguistic analysis of Czech sign language (especially its hierarchical structure), obtained through quantitative linguistic methods, will contribute to the development of the teaching theory and practice of Czech sign language. In terms of its concept and research methods, it is a seemingly unique research project which applies newly created quantitative linguistic methods to Czech sign language for Deaf people, which has (in contrast to spoken languages) a simultaneous and polysynthetic structure.

National sign languages of the Deaf have been considered natural languages by linguists for more than fifty years. Judging by the available professional publications it does not appear, however, that this conclusion could be verified in the light of the foreseeable validity of the laws of quantitative linguistics.

Our main task is therefore to apply and validate quantitative linguistic methods and analyse the relationships between language levels in sign language. A number of studies in this field have been already done, but from other perspectives (see, for example, Borneman et al., 2018; Handouyahia et al., 1999; Malaia, 2017; Malaia et al., 2016; Stewart, 2014; Uras, Verri, 1995). However, our tools and approaches are completely different from those applied in the studies mentioned above.

The single chapters of our books are organized as follows. The first chapter can serve as an introduction to a general study of sign languages. It is an English translation of the main part of the lecture notes (cf. Langer, 2013). The elements of quantitative linguistics, including historical remarks, have been collected in Chapter 2. This chapter was completed by the extension of related entries form the *New encyclopedic dictionary of Czech* (cf. Karlík, Nekula, Pleskalová et al., 2017). Chapter 3 deals with one of the most important laws of quantitative linguistics, namely the Menzerath–Altmann law. Its content was taken from the theoretical part of the paper Andres et al. (2012b). Language fractals are considered on this basis in Chapter 4, which was selected from the paper Andres (2014). Our methods are briefly sketched in Chapter 5, where a particular attention is paid

to a possible visualization of language fractals. Serious problems with the constituents of signs (especially measuring of their lengths) are also indicated there by means of the arguments from Langer et al. (2020) and Langer, Rypka (2017). The standard statistical criteria for the verification of obtained results are described in Chapter 6, which is based on the paper Andres et al. (2014). For the first experiment, treated in Chapter 7, where the validity of the Menzerath–Altmann law has been investigated for the sign language levels (both separately and together) via a fractal-based technique, we employed the results in Andres, Benešová, Langer (2019). For the second experiment, in Chapter 8, about the persistence of a sign language speech, explored by virtue of the Hurst formula, we adapted our study in Andres, Langer, Matlach (2020). The challenges for our future research in this field are indicated in Chapter 9. A possible profit of a quantitative linguistic analysis of Czech sign language, with a special respect to educational aspects, is discussed in Chapter 10, where the conclusions of the questionnaire in Langer, Rypka (2017) were taken into account. After an extensive bibliography, the profiles of authors and key figures have been supplied.

Olomouc, 2020   The authors

Peter Angermann, *A. I. (Nohandsfractal)*, 1989, silkscreen

1   SIGN LANGUAGES

1.1   Terminology

Hearing is one of the most important proximity sensors for humans, and apart from its signalling function, serving to inform us of sounds from the outside environment, it also has an important role in the perception of information presented verbally. *Deterioration in the function of the auditory analyser may result in hearing impairment, which limits the individual's interpersonal communication to a varying extent and creates numerous obstacles.* Unlike other types of impairment (e.g. physical, visual, or mental), hearing impairment is usually not physically apparent, which often causes misconceptions among the lay public related to its severity and the complications it causes to the person affected.

The development of communication competence in persons with hearing impairment differs significantly as a result of the sensory impairment. *The means of communication for the hearing majority of society and persons with less severe hearing impairment* (hard of hearing) *is the national language* (i.e. Czech in our environment), transmitted mostly in the form of spoken language (and the less frequently used written form of language), while *the natural language and means of communication for persons with more severe hearing impairment* (deaf) *is the national sign language* (e.g. the Czech sign language). The language specifics of persons with hearing impairment give rise to a certain cultural difference on the part of this minority, and make its socialization within the majority society more difficult.

From the perspective of special education, *the deaf are persons whose loss of hearing does not enable them to understand human speech even when compensated for with technical aids,* and who therefore depend on other means of communication in order to receive information. For the purpose of Act No. 155/1998 Coll., on communication systems of deaf and deafblind persons, sup-paragraph 2, the deaf are defined as persons "deaf from birth or having lost their hearing before acquiring speech, or persons suffering from complete or practical deafness having lost their hearing after acquiring speech, and persons severely hard of hearing whose extent and character of hearing impairment does not enable them to fully understand speech through hearing."

In the 1990s, part of the Czech community of hearing-impaired people followed the example of users of American sign language, and started labelling themselves as the Deaf (starting with a capital letter). People with hearing impairment affiliating themselves with the Deaf view themselves as a cultural and language minority and request not to be considered as impaired (cf. Hrubý, 1999). *The crucial uniting aspect within the community of the Deaf* is the specific communication system – *sign language*, which is significantly different from the communication systems used by the majority hearing society. In order to become a member of the Deaf community, a person has to fulfil numerous requirements, and most importantly, identify with the specifics of the culture and community of the Deaf (cf. Holubová, Motejzíková, 2002; Mrzílková, 1996; Lane et al., 1996). The Deaf therefore do not automatically associate all persons with hearing impairment, not even all persons suffering severe hearing impairment, or deaf persons (from the perspective of medicine or special education).

Visuo-motor communication systems play an important role in communication of hearing-impaired persons; these are systems using specific visual and motor means (shapes, positions, posture and movements of hands, posture of the head, mimics, etc.) to transfer information. There are, however, numerous terminology-related ambiguities leading to incorrect perception of the individual concepts and their content, and the subsequent incorrect utilization. One of the most frequent incorrectly used terms (especially by the lay public, but often also among the expert public) in relation to people with hearing impairment is the term *"znaková řeč"* [referring to sign language in English—literally "sign speech"].

Figure 1.1: A historical depiction of communication conducted in "sign speech" between two deaf men called *Conversation* created in 1925 by Leo Haas (1901 Opava – 1983 Berlin) (adapted from: http://www.zob.cz)

"The reason is that 'speech' is understood by linguistics as individual and unique acts of a person's utterances, i.e. as particular manifestations of language" (Macurová, 2001, p. 70), not as a comprehensive system of units (signs, words)

and rules for their interconnection into greater units—in this case, the term "language" is appropriate (see below). Similar cases would be expressions such as "Czech speech", "English speech" or "Russian speech"—ones which a literate person would never use to refer to the particular language system. The term "*sign speech*" has, however, been relatively firmly embedded in Czech terminology since the past, when it was not perceived from the linguistic perspective, but was used simply to refer to *a means of communication through visuo-motor means (i.e. a particular sign system)* regardless of their particular characteristics (evidence of this fact is e.g. Slovník znakové řeči [Sign Speech Vocabulary] published in 1988). From this perspective (it is in fact a *synonym for the term "signing"*), the utilization of "sign speech" is correct, but it is nevertheless better to avoid it due to this "lack of transparency".

Specialized research into communication of hearing-impaired persons, initiated in the 1990s, has shown that *it is necessary to distinguish clearly between the natural language of Czech deaf persons* (Czech sign language) *and the manually encoded Czech language* (signed Czech). Proponents of Act No. 155/1998 Coll., on communication systems were in all probability led to the utilization of "sign speech" by the need for establishing an *umbrella term* for two very distinct visuo-motor communication systems of hearing-impaired persons addressed in the Act: Czech sign language and signed Czech. In the supplementary bill developed by the Union of Deaf and Hard-of-Hearing Persons in 1997, the umbrella term "sign speech" is replaced with the term "znakový jazyk" [sign language]. This adjustment to the bill was not, however, included in the final version of the legislation, approved by the Parliament of the Czech Republic on 11 June 1998.

One of the inspirations for elaborating the above described bill was probably the Act of the National Council of the Slovak Republic passed on 26 June 1995 on sign language of deaf persons, which also used the analogical term "reč" [speech] (the term "sign speech" is also embedded in German terminology as "Gebärdensprache"). Although the original intention of the act was to ensure the right of people suffering severe hearing impairment to their own natural way of communication, the official legislatively embedded definition of "sign speech" as an umbrella term resulted in the *creation of "terminological chaos"* in the area of communication of hearing-impaired persons, as well as in their practical life: officially, there were only interpreters into "sign speech", TV broadcasts featured the news in "sign speech", etc. but a deaf spectator could not know in advance whether the TV news would be presented in signed Czech, or in Czech sign language, or if the interpreting would be conducted by an interpreter using the way of communication preferred by the client. For the reasons presented above, the term "sign speech" should be entirely eradicated from professional as well as lay terminology, and should be replaced with terms referring directly to the individual means of communication of persons with hearing impairment.

Macurová (1998) also points out the difficulties resulting from the utilization of the term "*znak* [sign]" as a noun or an adjective. From the linguistic (or more precisely semiotic) perspective, this term is understood as "something (a sound, an image, an object, an act, an event, ...) representing something else for some-

body (the transmitter and the receiver)" (Karlík, Nekula, Pleskalová et al., 2002, p. 567). The linguistically understood term "sign" is therefore structurally superordinate to the term "sign", referring to the lexical unit of a sign language.

In Slovak terminology, when referring to the lexical unit (or the visuo-motor communication system of hearing-impaired persons itself) in order to avoid this clash of meanings, the terms "*posunek, posunkový* [gesture—as a noun or adjective]" are used. While this eliminates the ambiguity of the term "znak [sign]", it induces another ambiguity—the two meanings of the term "posunek [gesture]". Although the term "posunek [gesture]" was generally used in the past to label the lexical unit, present-day Czech terminology distinguishes between "posunek [gesture]" (a synonym to a gesture) from "znak [sign]", because it is non-verbal in nature and impossible to be analysed (while a sign is verbal in nature and can be analysed as a lexical unit of language) (cf. Macurová, 1998). The term "znak [sign]" is used for reference to the lexical unit of the visuo-motor communication system of hearing-impaired persons analogically even in foreign-language resources ("sign", "signo").

For the above explained reasons, in accordance with current trends in Czech as well as foreign terminology, this text will use the terms *"sign language" and "Czech sign language"* (similarly to terms such as American sign language, British sign language, la langue des signes française, etc.) to refer to the natural visuo-motor communication system of hearing-impaired persons. The artificial visuo-motor communication system, using signs from Czech sign language as its basic lexical units and the grammar of the Czech language, will be referred to as *"signed Czech language"* or *"signed Czech"*.

1.2   Communication of deaf people

Verbal communication

A typical human way of communication is verbal communication, which means a transfer of information using language and speech; the basic unit of verbal communication is therefore a word as a lexical unit bearing meaning. De Saussure's concept (cf. De Saussure, 2008) of the relationship between language and speech distinguishes between:

- language as a system (la langue),
- speech as a general human ability (le langage), and
- utterance—utilization of language in communication (la parole).

Language (langue) refers to *a system of semiotic signs* (see Section 1.3 Natural sign systems – sign languages) serving for verbal communication of people. It follows a set of certain language norms valid in a particular community. Languages and their utilization follow strict semantic, grammatical, phonological and phonetical rules that apply only to the particular language belonging to a particular ethnic or an another culture-based group. According to the present-day concept of linguistics, all languages share certain common features (the so-called theory of universals), and *in order for a particular communication system to be regarded as a natural and full-fledged language, it has to include conventionally agreed upon*

basic attributes of natural languages (see Section 1.3.1 Sign language as natural language).

Within the internal structure of the language, it is possible to identify the individual language planes—the sub-systems of language characterized by the homogeneity of their units:

- *the lexical-semantic plane,* which involves an active and passive lexicon, i.e. the particular lexical units used for the individual semantic meanings;
- *the morphological-syntactic plane,* which involves the grammar and syntax of the language, i.e. the rules for correct grammatical construction of lexemes and their inflection;
- *the phonetical-phonological plane,* which involves the area of production and perception of the smallest units of lexemes that bear a phonological meaning—phonemes;
- *the pragmatical plane,* which involves the ability to use the language through speech (la langage) for social interaction through communication.

Speech (langage) is used in linguistics to refer to the way of utilization of a certain language for interpersonal communication and is regarded as a general human biological quality, where the transmitter transmits meaningful information to the receiver through acoustic, visual, tactile or other codes, and the receiver decodes and processes the information. *The utterance (parole)* is a unique and particular language manifestation produced by the communicant in a particular moment in a chosen communication code—it differs according to who, where, when, how, and under what conditions the utterance is produced.

As stated above, *the practical tool of verbal communication is speech (la langage); its means may be spoken or sign languages.* In the case of the majority hearing population, the basic means of communication is spoken speech and its acquisition in early childhood becomes the basis for the subsequent development of reading, writing and other education factors. Under favourable conditions, most members of society acquire the spoken language of their group rather easily within the first few years of their life. If there is, however, serious or long-term deprivation in any area important for language acquisition, it results in communication and language deficits. Hearing impairment and the consequent sensory deprivation is, depending on its extent, etiology and the time of its emergence, the cause of impaired and also limited development of the required communication competence in the majority language.

Nonverbal communication

An integral part of human communication interactions is nonverbal communication, which is an important complement of verbal communication but in some cases may also serve as an independent means of meaning transfer in a form different from language. The basic factor distinguishing nonverbal communication signals from verbal word expressions using language is that they have multiple meanings. Although every nonverbal signal has its general meaning, most of these signals cannot be unambiguously transferred into words through the process of decoding or interpreted precisely without a context (cf. Vybíral, 2005).

Strnadová (2008) remarks that some elements of nonverbal communication are innate in humans (e.g. laughter and cry), so they are instinctive, common to all people, and normally cannot be influenced intentionally. In contrast, other manifestations of nonverbal communication (or nonverbal behaviour) are conventional, learnt through observation and imitation of nonverbal expressions of other people in the individual's social environment. The ability to understand these signals of nonverbal communication is therefore conditioned by an acquaintance with the particular culture, social group and environment, as well as by previous experience. From the perspective of the receiver of expressions of nonverbal communication, the following types of nonverbal communication can be distinguished:

- individual, which does not primarily serve for communication, but occurs in every individual as part of interpersonal communication, being a manifestation of thinking (individual nonverbal communication is expressed by a person e.g. even if the person is alone in the room);
- interactional, which is intended to transfer certain meanings to other communicants.

According to Křivohlavý (1988), Volek (2002), Mareš and Křivohlavý (1995), elements of nonverbal communication, sometimes also referred to as "body language", are typically classified within eight basic types:

- mimics,
- gestures,
- posture,
- kinesics,
- eye contact,
- haptics,
- proxemics,
- physical appearance and environment.

According to Tarcsiová (2008), nonverbal communication plays an even *more important role* for deaf persons *than for hearing persons.* This is true due to the high dependence on the visual communication channel, which is closely related to nonverbal communication. While in interpersonal communication of intact persons, the above listed types of nonverbal communication tend to be independent from the simultaneously produced verbal language, with deaf persons, some of them (especially mimics, posture and kinesics) are an integral part of visuo-motor sign language, and contribute to grammatical phenomena produced when communicating through this language in the form of non-manual factors.

As stated above, a necessary prerequisite for any communication among people with hearing impairment is *intensive visual contact.* If visual contact is not established, effective communication cannot take place (in case of persons with less severe forms of hearing impairment, it is significantly limited). This dependence on visual contact during the stage of development of communication skills in hearing-impaired persons considerably promotes their ability to precisely distinguish and "read" visible nonverbal expressions. Concerning this fact, a crucial role in communication of persons with hearing impairment is played by *mimics*, and mimic expressions are manifested much more in them than in hearing persons

(the importance of mimics is so great that if communicating in the sign language, mimics even convey grammatical meanings). There are significant differences even in the area of *proxemics*, i.e. the mutual proximity of communicants. In general, the standard distance in the case of communication involving persons with hearing impairment is smaller compared to communication involving hearing persons. The reason is particularly keeping a certain optimal distance for lipreading (e.g. in case of hard-to-hear persons) and the possible need for mutual physical contact (*haptics*) which is necessary e.g. in order to establish visual contact. There are strict social rules, however, for physical contact of hearing-impaired persons, and these need to be followed (cf. Tarcsiová, 2008).

Natural and artificial sign systems

Visuo-motor communication systems used in communication, rearing and education of persons with hearing impairment can be divided into two groups concerning their origin. The first group are *natural sign systems*, i.e. those that were established through long-term natural development. These specifically involve national sign languages of deaf persons, and one of their ontogenetic stages—so-called children's sign languages. The second group are *artificially created sign systems*, which serve as "compromise" language systems enabling mutual intercultural communication of the deaf and the hearing, similarly to finger alphabets and sign language phoneme codes.

Artificially created sign systems

Unlike natural communication systems, artificially created communication systems did not emerge through spontaneous and long-term development, but as a result of an intentional and controlled process aimed at a particular goal. The category of artificially created sign systems involves in particular manually encoded spoken languages (e.g. signed Czech), Makaton and Gestuno.

Signed Czech (or signed Czech language) belongs to the group of manually encoded spoken languages. From the linguistic perspective, signed Czech is not a natural language, unlike Czech sign language; it is an artificial system created by hearing people to be able to communicate with deaf users of Czech sign language. A sentence articulated in signed Czech is constructed according to the grammatical and syntactic rules of the Czech language, but instead of words it uses signs "borrowed" from Czech sign language. Signed Czech is therefore a so-called pidgin—a combination of two independent language systems.

For hearing users, the acquisition of signed Czech is easier than learning Czech sign language because they do not need to learn an entire new grammar system; they only apply the features of Czech sign language to their existing knowledge of Czech. For deaf persons, however, it may not be so easy to master signed Czech—the reason is that it uses completely different grammar structures than their natural sign language. Utilization of singed Czech when communicating with a hearing-impaired communication partner is, nevertheless, beneficial particularly because the production of an articulated Czech sentence, concurrently

complemented with the signs of the Czech sign language, considerably facilitates lipreading (cf. Hrubý, 1999). It is therefore suitable especially for persons with hearing impairment who are competent in Czech, i.e. hard-of-hearing persons and post-lingually deaf persons (who may prefer this as the optimal way of communication).

Another artificially created sign system is *Makaton*, a language program using a system of manual signs and graphic symbols. It provides a basic means of communication and promotes the development of speech, as well as the comprehension of concepts in children and adults suffering communication issues. It was created in the UK in 1976 by a team of a speech therapist and two psychiatrists as part of a project focused primarily on enabling communication for adult deaf persons having difficulty learning new information and for hearing mentally disabled or autistic children having difficulty comprehending speech who did not communicate verbally (Janovcová, 2004; Tarcsiová, 2005). At present, Makaton is used especially in schools, hospitals, training and social-educational centres and homes for disabled children and adults as a tool for both communication with hearing-impaired persons and within alternative and augmentative communication.

The system comprises three components: a dictionary, signs and symbols. The principle of the system and the method for training Makaton are based on the signing itself (movement of one or both hands). It is always accompanied with speech, and to facilitate comprehension of the meaning, the content of the utterance is supplemented with mimics, voice modulation and the situational context. Usually there is no need to sign all the words constituting the sentence—only the so-called key words are signed, i.e. those bearing meanings important for the meaning of the entire message. In accordance with the individual development of the expressive component of speech and comprehension of the concepts, it is gradually possible to reduce the number of signed words. To further facilitate communication and comprehension, it is possible to supplement the individual signs with relevant symbols (utilization of multisensory approach).

Figure 1.2: The cover of a Gestuno dictionary published in 1975 (adapted from: www. amazon.co.uk)

Similarly to manually encoded spoken language (signed Czech), *Gestuno* is also an artificially created system, i.e. it is not a natural language. It was created as an international sign system intended especially for interpreting of official texts (cf. Krahulcová, 2002). Gestuno is an analogy to the international spoken

system—Esperanto. As with Esperanto, Gestuno did not experience any wider application in practice and is only used by a relatively small group of professionals.
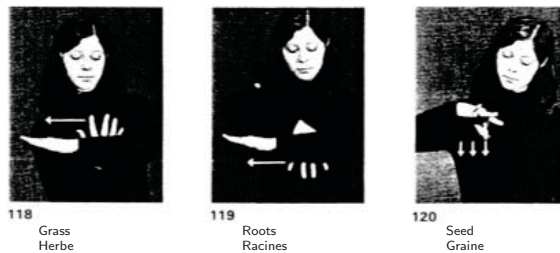
Figure 1.3: Gestuno signs (adapted from: http://inthelandofinventedlanguages.com)

Natural sign systems – Sign languages

The term "*sign language*" is used in specialized publications to refer generally to various *non-vocal language systems used by minority communities of the deaf of various nations*. It differs from vocal spoken languages especially in its nature (unlike with audio-oral spoken languages, sign languages are visual-motor in their nature).

Meaning in sign languages is not carried by sound, but instead by *two types of specialized* visuo-motor *media*:

1. lexical meanings are usually carried by manual media (the so-called manual factors—place of articulation, shape, position and movements of a hand/hands);
2. grammatical meanings are prevalently carried by non-manual media (the so-called non-manual factors – mimics, position and movements of head and torso).

Another important aspect in which sign languages differ from spoken languages is the *utilization of three-dimensional space*, especially in text (co-)reference, to express subject-object relationships and time expressions (Macurová in Karlík, Nekula, Pleskalová et al., 2002).

1.3   Sign language as a natural language

The perception of the importance of sign language has been significantly changing throughout human history; at the turn of the 1900s, there was even an approach stating that deaf persons could learn to speak more or less on their own, if lipreading spoken language and being prevented from any other form of communication—the only prerequisite would be to prohibit sign language as well as the use of natural gestures. The frequent *reasoning of opponents* of sign language, which has lingered on until the present, is that *sign languages are not full-fledged natural language systems*.

In 1960, the American linguist *William C. Stokoe* published his work *Sign Language Structure*, where he presented the results of a large-scale linguistic analy-

sis of American sign language (ASL) and demonstrated that *sign languages of the deaf have all the necessary features of natural languages* and are therefore *full-fledged languages*. These features especially involve the existence of double articulation (segmentation) in sign language, systematism, sign-based nature, productivity and the historical context of sign languages (cf. e.g. Černý, 1998; Karlík, Nekula, Pleskalová et al., 2002; Bímová, 2002; Slánská-Bímová, Okrouhlíková, 2008).

In general linguistics, a natural language is considered a "system of signs serving for human communication" (cf. Slánská-Bímová, Okrouhlíková, 2008, p. 13). Characteristic features are especially the possibility of (reproducible) communication between two partners or among more individuals, based on a system that is complex and dynamic and that enables a combination of its signs in accordance with certain rules (Čermák, 2001).

The basic properties of the semiotic sign are specifically *arbitrariness and duality* (cf. Karlík, Nekula, Pleskalová et al., 2002). Slánská-Bímová (in Slánská-Bímová, Okrouhlíková, 2008, p. 14) labels the semiotic sign as something (a medium, a form, a vehicle) representing something else based on the shared experience of the communicants. It can have the form of e.g. an acoustic signal (sound), a visual image (graphic representation), a referential object, an act, an event, a movement, etc. In the case of sign language, the form is visuo-motor expression (through manual and non-manual means). According to de Saussure, the semiotic sign is understood as the *signifiant (the signifier)*, while *a particular thing, object, quality or phenomenon that is communicated through the signifiant is the signifié (the signified)* (cf. De Saussure, 2008).

Based on the relationship between the signifiant and the signifié, Peirce (sec. cit. in Slánská-Bímová, Okrouhlíková, 2008) distinguishes between *three types of semiotic signs – icons, indexes, and symbols:*
- If the signifiant has an obvious relationship to the signifié (they are similar in their auditory or visual characteristics), the sign is *an icon*. Within spoken languages, these are usually interjections; in sign languages this incorporates the so-called visually motivated signs (imitative signs, see below).
- *Indexes* in natural languages (spoken as well as signed) usually represent various kinds of pronouns (personal, demonstrative, possessive), or adverbs. They are always directly tied to the entity of the transmitter producing them, and if used in the same lexical form by another transmitter, their meaning changes.
- *Symbols* are semiotic signs that have no relationship between the signifiant and signifié, i.e. they are entirely arbitrary and not motivated in any way. Symbols are more or less arbitrary successions of phonemes whose audio-oral (in spoken languages) or visuo-motor (in sign languages) expression is not similar to the meaning denoted.

According to Slánská (in Slánská-Bímová, Okrouhlíková, 2008), signifiants should have the following properties: they are accessible with senses, arbitrary, conventional, distinctive from other signs within the system, discreet, linear, semantic in nature, relocatable, culturally transferrable, stable, subject to prevarication, negativity and differentiality. *At present, linguists usually regard the in-*

dividual characteristics of semiotic signs to be fulfilled by sign languages, only based on perception with different senses; spoken languages use hearing (or sight when perceiving the graphic form of language), while sign languages use sight. Meier (2009) specifically emphasizes the fact that another sense used for perception of semiotic signs (i.e. even the language itself) may be touch; he refers to the work of David Quinto-Pozos, who points out the third modality of signs—the tactile-motor modality (referring to tactile sign languages for deafblind persons).

As mentioned by Čermák (2001), *according to the theory of universals, the presence of certain features is presumed in all languages.* Apart from general construction-related features of the language, this also involves certain formal and substantive universals (according to Chomsky's conception) and implicational universals. Although it is impossible to prove their presence in all languages existing and used throughout the world, the existence of particularly the following absolute universals is presumed:

- the presence of 10–70 phonemes, at least two of them being vowels (in the case of sign language, of a visuo-motor nature, the number of phonemes is much higher and their structure is different);
- utilization of deixis;
- utilization of means of quantification (expression of number with numerals);
- presence of relational (grammatical) words not possessing their own meaning, and of propria (proper names);
- sentences are constructed with two sentence members (subject–predicate);
- the order of attributes preceding a noun is always fixed (demonstrative pronoun–numeral–adjective).

According to Bímová (2002), the basic starting point for linguistic research into sign systems is *distinguishing between gestures and signs.* Stokoe proved that the sign (as the smallest lexical unit of sign language) may be, unlike the gesture, divided into smaller units—sign components (so-called phonemes). In his concept, every sign consists of three simultaneously occurring components: location in space, shape of hand/hands, and movement of hand/hands in space. In every particular sign language these components are "expressed through a qualitatively defined and quantitatively delimited set of items comparable to a set of phonemes within a spoken language" (cf. Bímová, 2002, p. 101). The possibility of dividing the smallest meaning-bearing units (signs) into smaller units not bearing the meaning but contributing to the distinction of the individual signs (sign components) brings one of the main attributes of natural language into sign languages: *double* articulation (phonetic-phonological segmentation).

Another necessary attribute of a natural language is *systematic nature:* every natural language is a stable system involving a set of otherwise independent discreet units and exactly defined relations between these units. Such a system has the following qualities: *hierarchy, anisotropy, and linearity.* Hierarchy in sign languages was proven with the already mentioned possibility of their phonetic-phonological segmentation. Anisotropy (different qualities depending on the direction) is a general quality conditioned by the way of human perception that is common for almost all man-made systems (including sign languages). The as-

sumption of linearity is not entirely fulfilled in sign languages, particularly due to their visuo-motor nature and the existence in three-dimensional space. Unlike spoken languages, sign languages primarily order their lexical units *simultaneously*. This property of sign languages is a result of the possibility to transmit and receive several pieces of information and units at the same time, while with spoken languages (that are audio-oral in nature) it is impossible to produce several phonemes at the same time. Even sign languages have, however, a fixed order of the individual items, thus the concept of "linearity" is usually replaced with the concept of "*sequentiality*" (cf. Bímová, 2002; Slánská-Bímová, Okrouhlíková, 2008).

Sign languages also fulfil another property of natural languages—they are *sign-based* in nature, i.e. there exists a system of signs (the concept of "sign" in this case is understood from the perspective of semiotics, i.e. as "something representing something else"—see above). In order for these "language signs" to fulfil their function, they need to be *accessible with the human senses*—sign language, perceived with sight, obviously fulfils this prerequisite. The individual language signs must also be *conventional*, distinguishable from other signs in the system, and *arbitrary*.

While the former two properties are fulfilled by sign languages, *the requirement for arbitrariness is not fulfilled completely*. The nature of the majority of signs (as lexical units) within Czech sign language is that of an icon (i.e. the "look" of the sign imitates the object it represents); arbitrary signs are in a minority. In contrast, with spoken languages the vast majority of language signs are arbitrary—the only type of signs that may be considered icons are interjections. Existing research into sign languages (including Czech sign language), however, show that their lexical units, i.e. signs, are not fixed; in contrast, they evolve over time in such as way as to become *arbitrary signs* (for instance, two-hand signs are transformed into single-hand ones, and the articulation space is reduced). In view of this, even this condition may be considered fulfilled by sign language.

Another property of a natural language is its *productivity and autonomy*. Productivity (the ability of a limited set of means to create an infinite number of combinations using a limited set of rules for their combination) of sign languages is ensured with the possibility of *combining the individual components of a sign into higher-level units*—particular signs (the phonetic-phonological level), and their subsequent combining into fluent messages (the morphological-lexical level). A prerequisite for the autonomy of a natural language is fulfilled by sign languages with their *ability to express the past, the present, and the future, a question, a condition, an imperative*, etc.

The last basic attribute of a natural language is its *historical context*, which is indisputable in the case of sign languages. The existing research into national sign languages has shown that ontogenetic development of sign languages takes place analogically to the ontogenetic development of spoken languages, and sign language are also subject to phylogenetic development.

Slánská-Bímová and Okrouhlíková (2008) also list other definitions of properties of natural languages that distinguish them from other communication systems

(according to the linguist Ch. F. Hockett):

1. the audio-oral nature of language; Hockett admits two exceptions—the written form of a language and sign language;
2. the possibility of transmission and directional reception of language signals;
3. temporariness and transience of language signals;
4. mutual interchangeability of the roles of the communication participants;
5. full feedback and the possibility of self-control;
6. specialization of articulation organs;
7. the semantic nature (language signs are linked to particular concepts);
8. arbitrariness of language signs;
9. discreetness of basic language units (phonemes) and the possibility to classify them within one of several clearly defined classes;
10. language autonomy;
11. language productivity;
12. transfer of traditions and cultural transferability;
13. double segmentation.

Additional characteristic properties of natural languages were gradually added (cf. Slánská-Bímová, Okrouhlíková, 2008):

14. the ability of a language to reflect upon itself;
15. the possibility of a language to be learnt even by a non-native speaker;
16. unlimited content (the possibility to express something novel through new signals);
17. independence from the stimulus (a language is not stimulated by either the outer environment, or the internal emotion);
18. contextual appropriateness (a language can be adapted to a new situation);
19. creativity (the ability of a language to create an infinite number of new sentences);
20. dependence on a structure (a language depends on the systematic structure of its components).

When concerned with the above listed properties of natural languages, it is obvious that sign languages and their features fulfil these characteristic properties (with the exception of the audio-oral nature of the language that was attributed to it by linguists), and *sign languages of the deaf have to therefore be considered natural languages* to which their users are naturally entitled.

In general linguistics, languages are usually classified into *language types*, and the principles of the individual typology theories are based on suitable conceptual topics, usually on the individual language levels. Čermák (2001) lists phonological, morphological, syntactic, and structural typologies. Phonological and morphological typologies are established according to the parameters of spoken languages, so sign languages cannot be classified with them due to their different form.

Syntactic (Greenberg's) typology classifies languages into six groups based on the word-order combination of the subject (S), the verb (V), and the object (O). According to Čermák (2001), the vast majority of world languages belong to the three most frequent combinations SVO, SOV, and VSO (approximately 85–90 %)

in total. Syntactic typology classification of Czech sign language has not been carried out, but research into other sign languages shows that *they usually fall into the categories SOV* (in sentences with transitive verbs) or SV (in sentences with intransitive verbs). Sentences with modal verbs manifested a different order of signs (cf. Massone, Curiel, 2004).

The last typology mentioned, structural typology (according to Skalička) is based on phenomena on multiple levels, and distinguishes between five types of languages: agglutinative, flective, isolating, introflective, and polysynthetic— while the so-called ergative (amorphous) languages are an entirely independent category (cf. Čermák, 2001; Černý, 1998). Servusová (2008) remarks that *sign languages belong to the group of polysynthetic languages (sometimes also called incorporating)*, i.e. languages that create complex lexemes (or signs in the case of sign languages) from the originally isolated lexical units, and these complex lexemes may contain subject, predicate, object, etc. and correspond to an entire sentence (cf. Černý, 1998).

In contrast to audio-oral spoken languages, which are based on the linear principle, the visuo-motor nature of sign languages enables the production of the individual language units (on the lexical as well as grammatical level) to be not only linear (the beginning and the end of a sign, or the production of signs one after another within a sentence), but also simultaneous. On the lexical level, simultaneity is manifested in the concurrent production of the manual (e.g. hand shape, palm orientation, movement) and non-manual (mimics, head movement, etc.) components of the particular sign. On the grammar level, this means the concurrent production of several signs through incorporation (see below), utilization of three-dimensional signing space, as well as the accompaniment with non-manual means.

The different forms of sign and spoken languages also influence the production speed of both these types of languages. According to Motejzíková (2007), research into foreign sign language has shown that although production of signs as complex lexical units is slower than production of words articulated aloud, *the production speed of a complex utterance in sign language is comparable to that of speech*, sometimes it may be even faster. This paradox is caused particularly by the possibility of simultaneous production of multiple meanings, especially through incorporation.

The individual national sign languages (their lexicon as well as grammatical structures) have evolved similarly to national spoken languages through long-term development influenced by many factors. *There is not one universal sign language* that would be identical for deaf people of various nations and cultures; however, thanks to their similar grammatical features, the individual sign languages are much closer to each other than the individual national sign languages to their respective national spoken languages. When communicating with each other, deaf users of various national sign languages are therefore able to rather easily come to a "compromising" *international contact sign system*. While the (sign) lexicon of communication partners usually differs, the basic grammar structure is normally very similar. The sign lexicon of the international sign system is naturally created

by the mutual adoption of signs and possible adjustments in order for the signs to be as comprehensible as possible for both communicating parties (cf. Tetauerová, 2008).

Manual factors in sign language

As already mentioned above, sign languages exist in the visuo-motor form, i.e. they are produced through specific movements of the human body. The main articulatory organ are arms, or hands—*all means of expression (semantic signs) produced with hands are called manual factors (bearers)*. The branch analysing the composition of the individual sign language signs is, as with spoken languages, phonetics and phonology.

The first linguistic phonetic-phonological study of sign language (particularly American sign language) was conducted by William Stokoe. In his treatise *Sign Language Structure* (1960) Stokoe demonstrated that American sign language has the "Saussurean" properties of a natural language, and also that *the individual signs* (unlike gestures) *may be divided into smaller segments–phonemes*. Due to the concern that terminological ambiguity may arise, Stokoe labelled these basic segments as *cheremes*, but later on, the phonetics of sign languages also adopted the term *phonemes*.

Stokoe originally distinguished between three basic parameters of the sign (a phonological unit):
- place of articulation;
- shape of the articulating hand(s);
- hand movements (cf. Macurová, 2008).

The number of phonemes within the individual parameters is different in various national sign languages, and unlike spoken languages where phonemes are produced in a linear manner, phonemes in sign languages are articulated simultaneously. Other linguistics-oriented research projects have shown, however, that the three above-stated segments of signs are not sufficient for a thorough analysis, and the number was therefore increased to the following currently used five parameters:
- place of articulation;
- shape of the articulating hand(s);
- position of hand(s) in relation to the body:
  - orientation of palm,
  - orientation of fingers,
- mutual position of hands;
- hand movements (cf. Macurová, 1996).

The place of articulation of a sign (tabula – TAB) determines all the parts of the articulation space that the hands reach when expressing the sign (for the standard size of the signing space, see above). The number of delimited sectors of the signing space determined depends especially on the thoroughness of analysis of this parameter; the location is usually recorded especially in the frontal plane of the body. The particular value for the parameter TAB is influenced by a number of factors, including the "canonical form" of the semantic sign, the context, the

influence of grammar (e.g. space visualization, incorporation, expression of tense) and other factors arising from the particular communicative situation (e.g. signing when sitting, walking, with one hand only, "whispering", etc.).

From the lay perspective, *the shape of signing hand(s) (designator – DEZ)* is the seemingly most important parameter of the sign, but the change in the individual variables within a fluent utterance usually leads to changes in the phonetical, not phonological plane. Again, the number of actually used variables (hand shapes) differs across the individual national sign languages. For Czech sign language, twelve basic hand shapes have been identified; these may be divided into five groups based on their relationship (cf. Souralová, 2004):

- closed hand;
- clawed hand;
- hand with fingers pulled to the palm;
- hand with fingers pushed away from the palm;
- closed hand with raised fingers.

As to the parameter position of hand(s) in relation to body (orientation), two independent features are distinguished – orientation of palm (ORI1) and orientation of fingers (ORI2). To determine the value of ORI1, the orientation of the inner wrist is important; for the value of ORI2, it is the orientation of the index finger, the middle finger, the ring finger, and the little finger in the upright position (the orientation of the thumb is therefore not so important). When the fingers are bent, the value is determined according to the orientation in the upright position. As with other parameters, the orientation of the palm and fingers may be different from the canonical form when actually producing the sign within a fluent utterance; this is caused especially by co-articulation of the individual signs.

The mutual position of hands (hand arrangement – HA) is a parameter observed in signs articulated with both hands. Research into Czech sign language determines the following mutual positions of hands (cf. Souralová, 2004):

- hands next to each other;
- one hand behind the other;
- one hand over the other;
- one hand inside the other;
- hands in mutual contact;
- hands joined together.

The last commonly observed parameter of the sign is *movement (signator – SIG) of hand(s)* articulating the sign. The visuo-motor form of the sign language enables making full use of the three-dimensional signing space, which makes this parameter a very variable category and which also makes its description difficult. Macurová (1996) distinguishes between the direction of movement (e.g. upwards, downwards, rightwards), the type of movement (circular, direct, swinging, etc.), its length (short, long, etc.), character (e.g. repeated, gentle, etc.), or possibly its absence.

A movement done when signing a particular sign may be further divided into the macro-movement of the entire arms, or forearms, and the micro-movement of the wrist or the individual fingers. Even within isolated signs, there is frequently

a combination of multiple movements that may be done linearly (one after another), or simultaneously (concurrently).

Based on the number of hands active in signing, their shape, and the type of movement, every sign may be classified within one of the following movement matrices (cf. Battison, 1978), based on its manual characteristics:

- Signs articulated with one hand:
  - signs with zero contact (e.g. STUDY, DAY, SUN),
  - signs with contact with the body, not with the other hand (e.g. DEAF, OLD, FRIEND).
- Signs articulated with both hands:
  - signs with both hands active and with the same shape (e.g. WEATHER, IMPORTANT, SLOVAKIA);
  - symmetrical signs where one hand is active (dominant hand, the right hand for right-handed persons), articulating above (under, behind, in front of, etc.) the passive hand; both hands have the same shape (e.g. SKIN, OCTOBER, DANGEROUS);
  - asymmetrical signs, where one hand (the dominant hand) is active, articulating above (under, behind, in front of, etc.) the passive hand; hands do not have the same shape (e.g. PRESENT, MACHINERY, NAIL).
- Composed signs combining the above listed types (e.g. MAY, LIBRARY, NOT BE ABLE TO).

Non-manual factors in sign language

While the above listed manual factors in sign languages are usually the bearer of lexical meanings, *the grammar component of the utterance is significantly represented through non-manual factors*, which are part of nonverbal communication. Nonverbal communication, manifested through non-manual bearers, has a considerable importance for the reception of information even in the communication of the hearing majority, but for persons with hearing impairment, who cannot use auditory feedback within communication interaction, its importance is considerably greater. Concerning their visuo-motor nature, non-manual factors are perceived with sight, and the perception takes place simultaneously with the perception of manual factors.

With regard to the place of production of non-manual means of expression, we can distinguish between *the movements and positions of the following body parts*:

- head (nodding, shaking, leaning, etc.);
- face and its parts:
  - eyebrows (raised, lowered, etc.);
  - eyes (viewing direction, degree of eye opening)
  - nose (neutral, or wrinkled);
  - cheeks (e.g. inflated, pulled in, tongue in cheek);
  - mouth (air flow, oral and speech components);
  - tongue (sticking out, licking lips, placed on teeth, etc.);
  - teeth (e.g. revealed teeth, lip biting);
  - chin (moving, retracted, etc.);

- shoulders (shrugging in various directions, movements, etc.);
- torso (e.g. leaning, rotation, bending);
- legs.

Oral components involve expressions manifested with the mouth that are not related to speech (e.g. lip vibration, yawning, one lip placed on the other), *speech components* are those movements of the mouth that are derived from the production of speech (de facto facial reflections of speech sounds—kinemes). Vysuček (2008) states that without oral or speech components, the individual signs would not make sense—they would be ungrammatical.

The absence of the auditory communication channel in persons with hearing impairment and their dependence on the visual perception of all motor (manual and non-manual) expressions transmitted by their communication partner usually results in significant development of the ability to differentiate and precisely analyse even subtle nonverbal expressions of a person that would be normally imperceptible for an intact person.

For hearing individuals trying to learn sign language, mastering of the appropriate passive as well as active utilization of non-manual factors is very often one of the most difficult stages in the study process. With regard to the fact that non-manual factors (as mentioned above) contribute significantly to the grammar component of sign language, the degree of their mastering usually determines the ability to sufficiently comprehend a signed message and express oneself fluently and clearly in sign language (inadequately used non-manual factors are usually the factor that enables the deaf to find out that the other communicant in sign language is a hearing person).

The focus of hearing-impaired persons (including those with less severe hearing loss, who do not use sign systems for communication) on visual perception of nonverbal expressions of their communication partner is extremely intensive, and it is one of the typical manifestations of a distinctive culture of persons with hearing impairment.

In sign language communication, non-manual factors fulfil the following roles: (cf. e.g. Souralová, 2004; Strnadová, 2008):

- *Expressing transmitter's emotions, intentions and attitudes,* including e.g. expression of the degree of urgency (through mimics, forward leaning of the body), subordination or superiority (mimics, position of shoulders), sadness or happiness (mimics, the overall energy of the motor expression), "shouting" or "whispering", etc.

- *Distinguishing between the lexical meanings of the individual signs,* especially semantic signs whose manual component is almost or entirely identical, so they may be undesirably mistaken (e.g. semantic signs FOG and DUST differ only in the degree of eye opening). Apart from non-manual factors, the elements significantly contributing to the differentiation of the individual semantic signs in case of signs with an identical manual component are even the sentence cotext and situational context of sign production.

- *Lexical representation of adjectives linked to signs for nouns*—apart from specific adjustment of the manual component of the sign (e.g. a different shape

of the hand or character of the movement—its extent or speed), non-manual factors are also commonly employed very significantly (e.g. inflated cheeks expressing greater size – BIG CABBAGE; tip of tongue between teeth – SLICED CHEESE; urgent facial expression – TERRIBLE HUNGER).

- *Gradation of adjectives and adverbs*—apart from the adjustment to the manual component of the sign representing a canonical form of an adjective (e.g. adjusted length of the movement trajectory or its speed), components added include mimic expression (e.g. BEAUTIFUL → THE MOST BEAUTIFUL – eyes and mouth open gradually more and more; SMALL → THE SMALLEST – mouth and eyes are closing, cheeks are pulled in; FAST (adverb) → THE FASTEST – head leans to the side, corners of the mouth are widening, the stream of exhaled breath grows faster). In addition to a very intensive non-manual expression, the superlative form is frequently accompanied by a manual classifier prefix, a semantic approximation of the meaning "the most". While gradation of adjectives usually involves three stages in traditional languages (the first – affirmative, from which the second – comparative, and the third – superlative are derived), in sign language it is impossible to determine the number of stages due to high variability and the possibility of employing various non-manual factors within sign production.

- *Modification of properties of some verbs,* where simultaneity of sign language is used, and this modification is incorporated in the canonical form of a sign together with its manual adjustment (in the case of spoken languages, the modification of the meaning of verbs is usually achieved through a combination with adverbs). Most frequently, verbs are modified by expressing the extent of effort or lack of difficulty needed to perform the particular activity (e.g. pedalling easily one's bike, learning hard).

- *Representation of adverbs themselves,* where non-manual components in combination with the context distinguish adverbs from the related noun (e.g. DISTANCE–DISTANT) and gradate it (see above).

- *Distinguishing the communicative function of the utterance,* i.e. the declarative, interrogative, imperative and wish function. The receiver of the information must be able to correctly decode the individual non-manual expressions in order to be able to react appropriately to the lexical content of the utterance. This is one of the reasons why the recipient of a signed utterance looks specifically at the transmitter's face while perceiving the manual signs with peripheral vision. The individual functions are commonly manifested through a stable set of non-manual expressions that are further adjusted or added to, as needed.

- *Expressing negation*—negation may be expressed simultaneously with the production of the manual component of a sign (usually through shaking one's head, a specific facial expression—mouth shaped into ∩, narrowed eyes), and may be combined with a particular manual sign indicating negation.

- *Expressing irony,* which is commonly expressed in spoken languages through voice intonation. In sign languages, the manual component of a sign is simultaneously complemented with a specific mimic expression (e.g. a grin).

Significance of sign language in the personality development of people with hearing impairment

The ability to receive information from our environment through hearing is vital to the development of speech and interpersonal communication and to the natural growth of mental and social abilities. Severe hearing impairment prevents the child from perceiving sounds, which leads to sensory deprivation, a limited variability of experience and impaired development of selected competences (cf. Koluchová et al., 1989; Gregory et al., 2001).

The most serious secondary disability related to hearing impairment is the *communication barrier*, which affects all aspects of personality development. Functioning and wholesome communication (notably in the case of a child with hearing impairment) should be the primary goal of those who are in contact with people with hearing impairment (family and teachers in particular). Potential emotional deprivation and *communication "isolation"* of the child may turn *sensory deprivation* into psychological deprivation, which in some cases has a negative impact on a previously harmonious family.

One of the ways to prevent a communication barrier is *ensuring optimum language development*, which is key to the cognitive and social development of a deaf child. Studies investigating how people with hearing impairment learn sign language, particularly American Sign Language (ASL) British Sign Language (BSL), have revealed that the *ontogenetic development* of sign language *is generally analogous to the ontogenetic development of spoken language* in hearing children (cf. Woll in Gregory et al., 2001). According to Evans (2001), however, the language development strategy aimed at communication in the most frequently spoken language of the given society in hearing children (see figure 1.4) is fundamentally different from the language development of deaf children due to the impaired audio-oral communication channel (see figure 1.5).
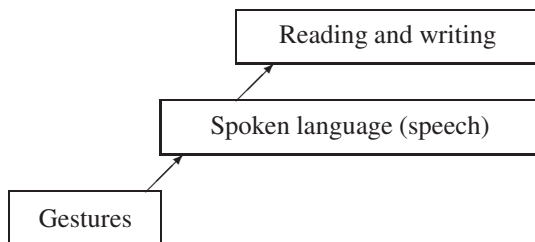
Figure 1.4: Language development in hearing children

While the natural language of hearing people is spoken language (Czech, in our case), children with hearing impairment are not predisposed to acquire it and are therefore unable to speak the same language code as their hearing community. As mentioned above, linguistic surveys (cf. Macurová, Homoláčová, Ptáček, 1997; Macurová, 2001; Evans, 2001) indicate that *sign language is a natural language for those with severe hearing loss and should not be denied to them*.

Regardless of which communication system they prefer, psycholinguists and experts dealing with the communication systems of people with hearing impair-

ment agree that *children with hearing impairment need to be able to access language as early as possible.* It is therefore necessary to create conditions and an environment for the child in the early period which would allow for successful communication between the deaf child and its mother. "Successful early communication is currently considered essential for language acquisition. Linguistic stimulation and functional feedback between the mother (and other family members) and the child in the early years are also fundamental to the healthy cognitive and emotional development of the child" (cf. Hronová, Motejzíková, 2002, p. 60). Evans (2001) reports that *evidence suggests that the features and specificities of sign language have prompted its early use as a medium for acquiring primary language competence.*
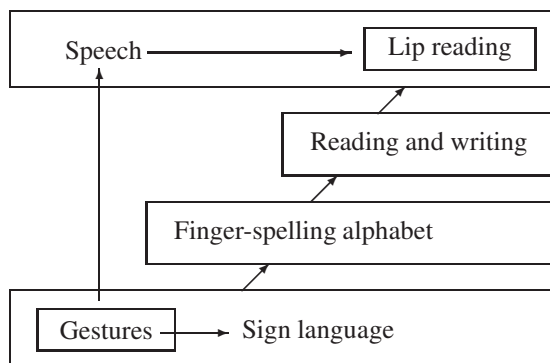
Figure 1.5: Language developmental sequence in deaf children

A study by Ahlgren (in Evans, 2001) also demonstrated that young deaf children exposed to sign language had advanced linguistic skills given their age (mainly because they used 'visual' sign language in the phase in which visual stimulation is important for linguistic and cognitive development). Another study by Ahlgren (in Hronová, Motejzíková, 2002) revealed that children with hearing impairment from deaf families who communicated with sign language were significantly more aware of their surroundings (continuously monitoring their surroundings in anticipation of new information) compared to deaf children from hearing families. Children raised using sign language therefore pay significantly higher attention to their immediate surroundings and learn randomly and spontaneously, thanks to which they learn to make the maximum use of their eyesight at an early age.

Svoboda et al. (2001) point out that the development of a child with hearing impairment largely depends on adequate education (to a much greater extent compared to intact peers), as deaf children are objectively limited in their efforts to generate incentives and information. The personality of a child with hearing impairment is substantially shaped by the environment in which the child grows up. More than 90 % of children with hearing impairment are born to hearing parents, for whom the birth of a deaf child tends to be a psychological and emotional challenge. One of *the most serious problems is the inability of parents to communicate with a deaf child naturally and fully due to the hearing loss.*

In contrast to deaf children of hearing parents, *deaf children growing up in a deaf family have better results and develop faster.* Originally, this was incorrectly attributed to the aetiology of their hearing loss (deaf children of deaf parents were expected to be less likely to have complex cerebral disability, while greater probability was linked with the hereditary causes of hearing loss). Research into this aspect of the development of children with hearing impairment (cf. e.g. Harris, 1978; Conrad, 1979 in Freeman et al., 1992; Evans, 2001) has shown, however, that "one of the positive factors which give deaf children of deaf parents an advantage is using sign language from the youngest age. There are, however, several factors. The attitude of deaf parents to their child's hearing impairment is, on average, incomparably more positive compared to that of hearing parents." (Freeman et al., 1992, p. 183) *Early communication in sign language is therefore undoubtedly beneficial for the linguistic, psychological and social development of the deaf child.*

Wholesome and effective communication plays a vital role in the comprehensive development of a school-age child with hearing impairment. Sovák (in Baláž et al., 1988, p. 70) defines the education process "as a specifically organized communication aimed at the transfer of knowledge, skills and habits. The primary requirement is that this communication be efficient and trouble-free." Krahulcová (2001, p. 69) also views communication as "the essence of the education of children with hearing impairment. Social and didactic communication supplies information, and the information then provides material for the development of cognitive processes at the highest levels of thought and speech, and in turn for the development of education in general."

Effective and "trouble-free" communication is objectively feasible *in the education* (psychological, cognitive and social development) *of children with severe hearing loss* only if visual-motor communication means (sign language and finger-spelling alphabet in particular) are involved. *Only sign language can be a means to rapid and accessible transfer of knowledge.* Sign language becomes the starting point and tool for developing other communication skills in children with hearing impairment (this is the basis of the principles of total and bilingual communication).

Sign language plays a crucial role in the personality development of a child with hearing impairment, and in *the search for identity* in particular. Given the hearing loss, deaf children are likely to meet and socialize with other deaf people in the future (the process begins the moment children start attending a school for the deaf). People with hearing impairment who meet specific requirements *consider themselves a cultural and linguistic minority* and refer to themselves as the *"Deaf" (capital "D").* Only those who have an *excellent command of the national sign language* become fully-functioning members of the Deaf community. People with hearing impairment, able to identify themselves with the specifics of the Deaf culture and enter the community, will most likely be able to establish social contacts, make friends and find partners. They will not be caught "halfway" between the worlds of the hearing and of the Deaf and will not lose their identity. This is sometimes the case with deaf or hearing-impaired people who have

received an oral education and are unable to communicate in a spoken language due to the degree of their hearing loss, and who have not learned sign language and are therefore unable to communicate with the Deaf either.

1.4   Types of sign language notation

The visuo-motor form of existence of sign languages makes them a suitable means of communication for persons with hearing impairment. Their form also provides them, however, with entirely different linguistic properties compared to audio-oral communication systems, derived particularly from the principle of simultaneity and the utilization of three-dimensional space. It is therefore very difficult for a native speaker of an audio-oral language (in some respects probably even impossible) to acquire all the lexical and grammatical specifics of these systems. The notation of visuo-motor sign systems is also a specific issue (especially regarding their linguistic analysis), since their simultaneity does not allow for utilization of standard writing as we know it from linear languages. For this reason, linguistic research focused on sign systems has gradually proposed several specific notation systems enabling their linear or graphic notation and allowing for quantification of the occurrence of individual phonemes.

An utterance of a human being expressed with visuo-motor sign systems (commonly referred to as signing, or sign speech) is, as with an utterance expressed with audio-oral systems (speaking, or speech), an unreproducible act.

To enable scientific investigation of the individual communication or language systems *there is a need to record these systems appropriately,* and subsequently to describe them using relevant techniques and methods.

To perform linguistic investigation of audio-oral systems and phenomena occurring in them, the analysis of their graphic form is sufficient in the majority of cases; however, in the case of sign systems, which are visuo-motor in their nature, such analysis is much more difficult (and in some places it has not been conducted yet). The difficulty could be compared to linguistic analysis of audio-oral systems where its permanent graphic form or a reproducible auditory form would not be available.

The most phonetically and phonologically universal option is *transcription with notation systems* where the sign manifestation, or parameters of the individual signs, are captured with specific codified tags and symbols typical for the individual notation systems. Signs recorded with notation systems are not recorded from the semiotic perspective, so it is impossible to analyse syntactic and contextual links from them. *They enable,* however, very detailed *analysis of the individual parameters* constituting the internal structure of the sign. The phonological and phonetic principles of the notation of these parameters also *enables statistical analysis* taking into account e.g. the frequency of occurrence of the individual variables, searching for the minimal pairs of signs, etc. The disadvantages of these notation systems are that learning them is quite demanding, and they are significantly arbitrary.

In a period when sign systems are globally investigated from the perspective of phonetics and phonology, dozens of transcription systems have been developed.

With regard to Czech sign language, we consider the following transcription systems the most important:

- Stokoe's notation system and its modifications,
- the Hamburg notation system,
- SignWriting,
- the Edinburgh non-manual coding system,
- the Berkeley transcription system.

Stokoe's notation system and its modifications

The original *Stokoe's notation system (SNS)* is based on defined manual parameters of signs (see the sub-section 1.3 Manual factors in sign language) and has become the source of many other revisited systems, modifications, and adjustments for the individual national sign languages. *The notation system for Czech sign language* was designed in 1996 by Alena Macurová in a proposal published in the journal Speciální pedagogika [Special Education] (cf. Macurová, 1996) and subsequently amended by Karel Beneš (cf. Okrouhlíková, 2008, 2012). It is based primarily on Stokoe's notation system (for American sign language) and its modifications from 1984 carried out by a team of British researchers Mary Brennan, Martin D. Colvill and Lilian K. Lawson (for British sign language) (cf. Okrouhlíková, 2008).

The notation system for Czech sign language focuses on its manual component from the phonetic perspective, and the repertory of phonemes is determined according to the theory of minimal pairs—pairs of signs of different meaning that differ from each other with only one phoneme. This phoneme has a phonological validity. As stated in the section 1.3 Manual factors in sign language, the individual phonemes may be divided into five groups of parameters: tabula (TAB), designator (DEZ), orientation (ORI), hand arrangement (HA) and signator (SIG). All signs may be divided into six types (see the same section) and recorded according to three basic matrices: signs articulated with one hand, signs articulated with both hands, composed signs (Brennan, 1990).

Notation of the individual phonemes within the superordinate parameters uses special symbols that are usually pictograms or ideograms. In the version designed by Macurová, there are thirty basic symbols for the parameter TAB (and nineteen more for the manual TAB). The place of articulation is understood as the place where a particular sign starts.

The symbols for parameter DEZ are based on the letters used in the Czech finger alphabet, and are complemented with special tags, indexes, and additional symbols capturing their slight modifications. Macurová determined twelve basic handshapes, and thirty more shapes derived from them. The parameter ORI is noted separately for the orientation of palm (ORI1) and for the orientation of fingers (ORI2) in relation to the signing person's body. The standard number of symbols used to capture orientation is six.

For signs articulated with both hands, their mutual position is also noted (HA). There are five symbols available; these are used in case the mutual position of the hands is not obvious from the notation of other parameters. The parameter

that is most difficult to note is undoubtedly movement (SIG). For its notation, there are 34 symbols capturing basic movements and twelve combined symbols for circular movements. There is also a special symbol for capturing the contact of a hand with another body part, or with the other hand (cf. Okrouhlíková, 2012).

In the notation system for Czech sign language, the individual symbols are recorded according to a strictly determined syntax, which is governed primarily by the movement matrix of the particular sign (based on the premise that the right hand is the dominant one):

- signs articulated with one hand—TAB DEZ $_{ORI1\ ORI2}$ ^{SIG};
- signs articulated with both hands, both hands active, both in the same shape—TAB DEZ $_{ORI\ 1\ ORI2\ (HA)}$ DEZ $_{ORI1\ ORI2}$ ^{SIG} (parameters for the left hand are recorded first, followed by those for the right hand);
- signs articulated with both hands, one hand is active, the other passive—manual TAB $_{ORI\ 1\ ORI2\ (HA)}$ DEZ $_{ORI1\ ORI2}$ ^{SIG} (parameters for the passive hand are recorded first, followed by those for the active hand);
- composed signs—the individual parts are recorded separately, separated with two diagonal lines (cf. Macurová, 1996).

Although all the recorded parameters are manifested simultaneously, the notation is linear. As mentioned above, the parameter whose notation is the most difficult is movement of hand(s), as it may be a combination of several isolated movements. If the individual movements are performed simultaneously, the respective symbols are noted vertically one under another (or horizontally in round brackets); if the individual movements follow one another, the symbols are noted horizontally next to each other.

In 2003, Karel Beneš modified the system designed by Macurová. The modifications concerned specifically correction of certain deficiencies in the original systems by introducing additional phonemes within the individual parameters. Beneš increased the number of the basic types of signs for the Czech sign language to seven, and they may be recorded in five matrices:

- signs articulated with one hand—TAB DEZ $_{ORI1\ ORI2}$ ^{SIG};
- signs articulated with both hands, both active hands are in the same shape—TAB DEZ $_{ORI\ 1\ ORI2\ (HA)}$ DEZ $_{ORI1\ ORI2}$ ^{SIG} (parameters for the left hand are recorded first, followed by those for the right hand);
- signs articulated with both hands, one hand is active (in the same or a different shape)—TAB DEZ $_{ORI\ 1\ ORI2\ HA}$ DEZ $_{ORI1\ ORI2}$ ^{SIG} (parameters for the passive hand are recorded first, followed by those for the active hand);
- signs articulated with both hands, both hands are active and in different signing spaces—TAB DEZ $_{ORI1\ ORI2\ HA}$ TAB DEZ $_{ORI1\ ORI2}$ ^{SIG} (parameters for the left hand are recorded first, followed by those for the right hand);
- signs articulated with both hands, each of them performing a different movement in a different signing space—TAB DEZ$_{ORI1\ ORI2}$$^{SIG}_{HA}$ TAB DEZ$_{ORI1\ ORI2}$ ^{SIG} (parameters for the left hand are recorded first, followed by those for the right hand);
- composed signs—the individual parts are recorded separately, separated with two diagonal lines (cf. Okrouhlíková, 2012).

As a result of the increased number of distinguishing phonemes, the number of symbols used for the individual parameters also increased. The number of basic symbols for the parameter TAB was increased to 62. The system of twelve basic handshapes (DEZ), derived from the Czech finger alphabet, was preserved, but the number of derived shapes was increased to 39. A significant increase in the number of phonemes may be observed for parameters ORI1 and ORI2, where the number was increased from six to 26, which enabled capturing subtle differences in hand orientation. In addition, the number of phonemes for parameter HA was significantly increased to 26. As with Macurová's system, contact of both hands is recorded with a special symbol, and the location of the contact is specified with one of 25 symbols noted in square brackets (manual TAB).
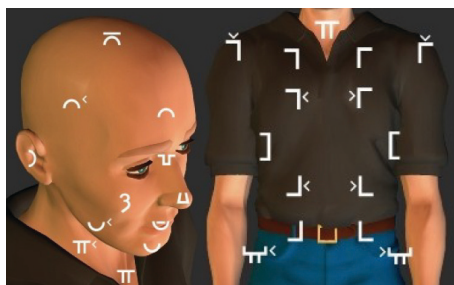
Figure 1.6: Symbols for notation of sign localization (parameter TAB) (adapted from: Okrouhlíková, 2012)
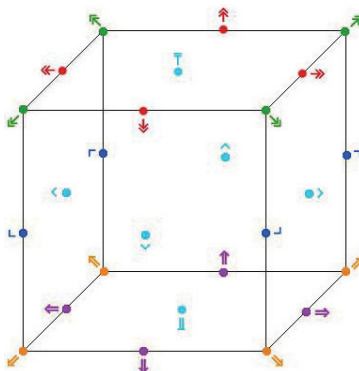
Figure 1.7: Symbols for notation of the orientation of palm and fingers (adapted from: Okrouhlíková, 2012)

A significant extension was also done for the notation of parameter SIG; the modification distinguishes between a direct movement (ORI1 and ORI2 remain the same; 26 symbols), a circular movement (ORI1 and ORI2 remain the same; 3 symbols; it is also possible to express the plane of the circular movement with indexes x, y, z), an arched movement (ORI1 and/or ORI2 change; 2 symbols), a wrist movement (ORI1 and/or ORI2 change; 4 symbols), a finger movement (the shape of hand changes; 6 symbols; in case of a complete change in the hand-shape, the new shape is noted in square brackets), and a movement of both hands

(5 symbols combined with symbols for a direct movement). Greater precision of the movement notation may be achieved with nine complementary symbols capturing possible modulations and modifications of movement (e.g. an alternating movement, a rapid movement). In contrast to Macurová's system, the individual movements performed simultaneously are noted horizontally in round or curly brackets; movements following one another are noted horizontally next to each other, and may be separated with a slash (cf. Okrouhlíková, 2012).
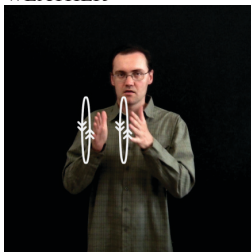
Figure 1.8: Comparison of the notation of sign parameters according to the system designed by Macurová (on the left) and the system modified by Beneš (on the right) (notations adapted from: Okrouhlíková, 2012)

Hamburg notation system

As with notation systems for Czech sign language (designed by Macurová and Beneš), *the Hamburg notation system (HamNoSys)* may also be considered a derivation from Stokoe's notation system; it is a phonetic transcription system recording the shape and orientation of the hand, the place of articulation, the movements of hands, and the movements of the head and upper body. Compared to the above described systems, the Hamburg system is more universal, since it uses icon marks to indicate the handshape, so it is not tied to a particular finger alphabet. It was designed in 1987 in Hamburg by a team led by Siegmund Prilwitz at the Institute of German Sign Language and Communication of the Deaf, and has been further developed (at present, version 4.0 is available) (Universität Hamburg, 2020). The advantage of this system is easy localization according to the conditions of any sign language in the world, i.e. even Czech sign language, or tactile language.

Unlike the above presented notation systems, *HamNoSys focuses on both the manual and the non-manual component* of the particular sign language. As to recording of the manual component of signs, the individual signs are divided into three basic types from which the transcription matrices are derived, having the specific syntax of the individual parameters and their symbols:

- single-handed signs: DEZ $_{ORI2\ ORI1}$ TAB ^{SIG};
- two-handed symmetrical signs: ^{HA} DEZ (dominant hand) $_{ORI2\ ORI1}$ TAB ^{SIG};
- two-handed asymmetrical signs: ^{HA} DEZ (dominant hand) $_{ORI2\ ORI1}$ DEZ (non-dominant hand) $_{ORI2\ ORI1}$ TAB ^{SIG}.

As stated above, in order to describe the handshape (DEZ), HamNoSys uses 31 iconic symbols in total. These are based on six basic and six derived handshapes and are combined and complemented with diacritic symbols representing e.g. the position of the thumb, the distance between the thumb and other fingers, bending of the individual joints of the hand, contact between fingers, etc. In total, the last version of HamNoSys (version 4.0) presents combined symbols for 150 various handshapes (Universität Hamburg, 2020). Orientation of fingers (ORI2—in total 18 symbols) as well as orientation of the palm (ORI1—in total 8 symbols) are recorded with a subscript following the symbol for the handshape, and may be complemented with some of the four symbols representing the bent wrist.

The place of articulation is only indicated if the sign is performed at a specific part of the body or space (when articulating in the neutral signing space, it is not indicated). There are 42 basic reference points of the human body thus far; these are further specified when combined with additional symbols (nine in total) informing whether the place of articulation is on the right/left side of the reference place, to the right/left of it, on the palm or the back of the hand, behind the body, in contact with the body, near the body or far from it. In total, it is possible to precisely specify more than 160 places or articulation within the signing space.

As with the previously described notation systems, HamNoSys uses a wide range of symbols and their combination for notation of the individual types of movement (parameter SIG) and their detailed characterization. There is a difference in the syntax of notation of absolute movements informing about the precise place of the beginning and ending of the movement (TAB ^{SIG} TAB), and of relative movements where the place of the beginning is defined followed by a combination of symbols referring to the movement that describe the movement into the space (TAB ^{SIG}). In the latest fourth revision of HamNoSys, the individual movements are classified as:

- path movements and their sub-types:
  - a direct movement with a specified direction and possibly its extent (small/big);
  - an arched movement with a specified basic direction and orientation of the arch;
  - a wavy or wiggly movement with a specified basic direction and character of the movement (wavy/wiggly);
  - a circular movement or its derivations with a specified direction of the movement, the size of the circle, the place of the beginning and ending of the

movement, partial circular movements, spiral movements or elliptic movements;

- local movements – from the macro-perspective, they do not employ the entire arms:
  - a change in the orientation or position of the hand;
  - finger movements (crumbling or flickering);
  - movements of the wrist or forearm;
- non-manual movements of the body, shoulders, head, etc. (notation uses the same symbols as those for other movements, combined with the parameter TAB).

WEATHER

SIGN

Figure 1.9: Notation of sign parameters according to HamNoSys (notations adapted from: Okrouhlíková, 2012)

To note down the character and properties of the individual movements, there are approximately 42 symbols, usually in the form of arrows, or arrows combined with other graphic figures (circles, arches, tildes, ellipses, etc.), as well as approximately thirteen additional symbols describing the quality of the movement (speed, tension, an abrupt halt, repetition, a change in hand dominance, etc.). The individual symbols as well as their combinations are recorded in a linear manner. If they are performed consecutively, they are noted in round or double angle brackets. If they are performed simultaneously, they are noted in square brackets. If the individual movements immediately follow one another and therefore merge, the respective symbols are noted in angle brackets.

The mutual position of hands (parameter HA) is recorded only for two-hand signs, and with regard to syntax, it is placed at the beginning of the entire notation. There are eight symbols used to represent various options of the mutual position of the hands.

Notation of the non-manual component of a sign language is not completely developed in the HamNoSys, and as stated above, non-manual movements may be recorded as a combination of symbols for movement and symbols for the place of articulation. While it is possible to record e.g. movements of eyebrows, eyes etc., the existing set of symbols for movement and place of articulation cannot capture all the currently used mimic expressions, and oral and speech components.

All three notation systems presented above record the individual components of signs in a linear manner, and when some phonemes are articulated simultaneously, this fact is indicated with special marks (e.g. placed in a certain type of brackets).

SignWriting

Another notation system, and a relatively frequently used one—*SignWriting*—records the individual phonemes with fixed graphic symbols into complex figures capturing the upper body of the signing transmitter. These complex figures therefore create a notation in a simultaneous manner, i.e. in the same way the articulation of the individual components is performed. As with HamNoSys, SignWriting enables the recording of the basic characteristics of the non-manual sign component. Unlike the other systems, however, which are primarily focused on the lexical level of the sign language due to their scientific-research utilization, SignWriting aims to become a full-fledged graphic representation of sign languages up to the level of sentences and even the entire messages; also capturing the syntax of the respective sign language. The individual lexemes (or their records in the form of complex figures) are (somewhat unusually for the users of standard graphic phonetic systems) arranged vertically into columns, so they look similar to the original sign writing systems (e.g. Chinese or Japan) (www.signwriting.org).

The SignWriting system was created in 1974 in Denmark by Valerie Sutton, and its principle is based on the system of DanceWriting used for the notation of ballet dance, designed by the same author. *SignWriting is not primarily intended for scientific purposes,* nor for phonetic description of the lexical component of signs; its aim is to be comprehensible to the lay public (the tags are in the form of icons) and widely accessible to its users for notation of national sign languages. For this reason, *it may become a commonly used script for sign languages.* While is differs from the other systems with its different principle of graphic combination of the individual symbols, a shared feature is the manner of analysis of signs into the individual parameters and their subsequent notation. In SignWriting, the following parameters are described:
- shape and orientation of hand,
- place of articulation,
- movement,
- contact,
- non-manual components (cf. Okrouhlíková, 2012).

The parameter notation principle in the SignWriting system lies in simultaneous graphic depiction; unlike the above described notation systems using linear notation, the system does not distinguish between the individual movement matrices and does not involve different patterns of parameter notation. The manner of

parameter notation is therefore based on the visual image of the signing transmitter, i.e. it captures the components of the particular sign that are necessary for its subsequent identification in a spatial arrangement (location of the individual symbols) corresponding to the reality from the perspective of the signing transmitter.

WEATHER
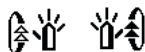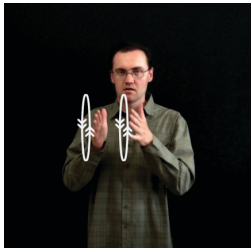
SIGN

Figure 1.10: Notation of sign parameters according to SignWriting (notations adapted from: Okrouhlíková, 2012)

Symbols for handshapes are dominant and initial in SignWriting; they are derived from three basic handshapes (closed fist, open fist and flat hand). With these three basic handshapes, the system distinguishes between ten basic sets of handshapes based on the number of fingers stretched out; the number of symbols for handshapes derived and classified within these ten sets is 261. *Hand orientation (of both the palm and fingers)* is captured according to fixed rules. The utilization of the symbol for handshape in the original form indicates the vertical orientation of the palm (parallel to walls), the degree of colouring refers to its orientation to the back (not coloured), to the front (coloured) or to the side against the central axis of the body (half-coloured, with the coloured part referring to the back of the hand). The orientation of fingers is indicated with various orientation of symbols for the handshape and palm orientation.

The place of articulation is captured in SignWriting with the above-mentioned symbols for the shape and orientation of hand which are placed in a particular place in relation to the reference point, usually the symbol for the head, or other body parts (the nape, the neck, shoulders and hips). If a sign is articulated in a neutral signing space, the reference point does not need to be indicated. The left-right orientation of the sign's location is encoded from the perspective of the producer of the sign (www.signwriting.org).

The parameter of movement, which is the parameter of a sign that is traditionally most difficult to record due to its three-dimensional and dynamic character,

is represented in SignWriting with special arrows (according to the type of movement; empty arrows for the left hand and full arrows for the right hand). The system distinguishes between:

- arm movements:
  - vertical movements (arrows with a doubled base line)—24 symbols;
  - horizontal (arrows with a single base line)—28 symbols;
  - horizontal-vertical movements (circular or arched movements)—11 symbols;
- forearm rotation movements:
  - vertically oriented (4 symbols);
  - horizontally oriented, directed away from the body (4 symbols);
  - horizontally oriented (4 symbols);
- wrist movements:
  - swinging (10 symbols);
  - rotation (vertical, horizontal, horizontal-vertical) (6 symbols);
- finger movements (12 symbols). (cf. Okrouhlíková, 2012)

In order to specify the movement more precisely, there are additional special symbols (nine in total) capturing its dynamics (fast, rapid, tension, slow, simultaneous movement of both hands, etc.). If the handshape changes in the course of a movement, this is indicated by adding the symbol for the final (new) handshape after the arrow indicating the direction of the particular movement.

Mutual hand contact, or contact between a hand and another body part is recorded within the complex of symbols at the exact place in the signing space where it occurs. The system distinguishes between six types of contact (a touch, permanent contact with a circular movement, short-term contact, holding, hitting and insertion), each of them having its own symbol (cf. Okrouhlíková, 2012; www.signwriting.org).

The notation of the non-manual component of signs is possible to a limited extent even in the above described Hamburg notation system, but this integral part of sign language communication, which considerably contributes to expression of grammar, is captured much more thoroughly in SignWriting. Special symbols placed into a circle symbolising the head refer to:

- shapes of eyebrows (6 symbols);
- shapes of eyes (6 symbols);
- viewing direction (8 symbols for the vertical direction and 7 symbols for the horizontal direction);
- appearance of cheeks (6 symbols);
- airflow (2 symbols);
- shapes of nose (2 symbols);
- speech component (16 symbols);
- oral component (21 symbols);
- expressions of emotions (10 symbols).

Movements of the entire head are expressed with arrows similar to those capturing hand movements (distinguishing between vertical and horizontal movements) and there are 16 symbols for them. As stated above, the SignWriting system is intended to become a full-fledged graphic representation of sign lan-

guages; for this reason, it also provides punctuation marks (in total, there are eight symbols for a period, a comma, a colon, a question mark, an exclamation mark, quotation marks, brackets, and a hyphen) that may be inserted into a continuous sign message (www.signwriting.org).

The Edinburgh non-manual coding system

A British research team (partly identical to the team that extended and modified Stokoe's notation system in 1984) developed *a notation system for recording the non-manual component of sign language signs* in 1982. In accordance with the home of the research team, this notation system was called "The Edinburgh Non-Manual Coding System". It was designed primarily to identify the meaning of the non-manual component of signs. The designers of the Edinburgh non-manual coding system took inspiration from two of the above listed notation systems—Stokoe's notation (from which they took the individual symbols) and SignWriting (whose principle of creating combined graphic shapes they adopted) (cf. Okrouhlíková, 2008).
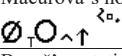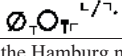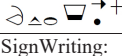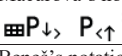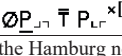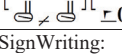
Figure 1.11: A comparison of sign notation according to the individual notation systems

As with SignWriting, *the transmitter's head is represented with* a circle, with the individual symbols indicating the movements or appearance of the individual parts of the face placed into or around it. (In contrast to SignWriting, the right-left orientation is recorded from the perspective of the recipient.) One dif-

ference from SignWriting is that the Edinburgh Non-Manual Coding System also involves recording of movements of other body parts, especially the shoulders and the torso. There are 18 basic symbols based more or less on Stokoe's notation system, or symbols with a slightly shifted meaning (14 of them describe the movement and four represent a particular position or shape of the respective part of the face). In addition, there are four additional symbols indicating other details of the movement (rapid, short, the number of repetitions).

Based on the position of the individual symbols within or around the circle, or with the symbols for shoulders and the torso, it is subsequently obvious which part of the face or body performs the indicated movement or is in a particular position. With some non-manual expressions (e.g. movements of mouth or eyes), the symbols for movement must always be combined with a symbol referring to the shape of the particular part of the face. If recording the non-manual component for isolated signs, it is recorded above the notation of the manual component (frequently with Stokoe's notation system). When recording the non-manual component of a continuous message (at the level of entire sentences), the notation is provided if a change occurs in the non-manual expression (so it is not necessary to indicate it for every isolated manual sign). Oral and speech components of signs cannot be recorded with the Edinburgh Non-Manual Coding System; there is a need to use written notes (cf. Colville, 1986; Okrouhlíková, 2008).

The Berkeley transcription system

An entirely different view of sign language notation is adopted by the last one on our list of common notation systems—*the Berkeley transcription system.* It was developed by American and Dutch scientists as an output of the project Berkeley Sign Language Acquisition Project investigated in 1998–2001; the aim of the project was to develop a way to record the interaction of deaf signing children with deaf and hearing communication partners, and to analyse it (cf. Hoiting, Slobin, 2002).

In contrast to all the above described notation systems, which capture the sign language and its individual signs from the perspective of phonetics and phonology, *the Berkeley transcription system focuses on the morphological and the semantic level, i.e. the level of meaning units* (cf. Slobin et al., 2001a; Okrouhlíková, 2012). A standardized form of PC notation and a set of software applications for analysis of the transcribed material is an integral part of the system. Regarding the fact that the Berkeley system focuses on the notation of complete communication interactions, it records not only the individual signs (their manual and non-manual components), but also the gestures and other manifestations of communication behaviour expressed. In order to record the individual observed parameters and phenomena, there is a set of symbols that are in fact English abbreviations of these elements, and in the case of national localization, they may be adjusted to correspond to the particular majority national language.

The basic building blocks of this transcription system are meaning units produced within a communicative act, thus recording them is also the basis of the

entire notation. The system uses notes with a semantic meaning in the particular majority national language complemented with other symbols indicating the respective manual and non-manual expressions, gestures, etc. The notation therefore enables the recipient to understand the content of communication, but it does not enable its reproduction without a knowledge of the particular sign language. Analysis of the nature of the observed manifestations of communication, modification of the individual signs, and the entire communicative behaviour subsequently allow for recognition of certain grammatical phenomena (cf. Slobin et al., 2001a).

Notation according to the principle of the Berkeley transcription system is based on observation of the following variables:
- individual signs (including equivalents) and their specification;
- signs for numerals;
- signs for demonstrative and possessive pronouns;
- poly-component signs (verbs);
- notes for expressions different from a sign in their nature (including equivalents);
- movement trajectories:
  - the shape of a movement;
  - the vertical direction;
  - horizontal sagittal direction;
  - horizontal lateral movement;
  - horizontal right-left direction;
  - swinging movement;
  - other directions;
- classifier handshapes (or specifiers) and the mutual position of hands:
  - classifiers for objects;
  - classifiers of holding;
  - specifiers of the shape of surfaces (the outer shape);
  - utilization of a body part instead of a particular sign;
- spatial relations;
- handshape (specific shapes);
- hand orientation (palm and fingertips);
- movement patterns (types of movement);
- non-manual components;
- the character of the movement;
- simultaneity of units (manual as well as non-manual);
- changes in handshape;
- extra-linguistic communicative behaviour:
  - gestures;
  - attention-drawing means;
- expressive and contextual situations:
  - mistakes and unconventional signs;
  - answers without signing;
  - continuous succession of utterances;

– interruption;
– repetition of an utterance by a partner;
– repetition and densification of an utterance;
● additional communicative aspects (attention and focus of communicants, oral communication, phonological description of gestures, etc.) (cf. Slobin et al., 2001b).

The individual signs and abbreviations used are designed so as to be available in the common sign set of a keyboard (ASCII) and are recorded in a linear manner into a line (with every shift in the communicative roles, a new line starts). The notation uses standard punctuation marks to separate the individual utterances.

An example: A transcription of communication in American sign language between a mother and her daughter (21 months old) (cf. Slobin et al., 2001b, p. 28).

*MOT: t@ag(*2) w@ag ˆopr'WHQ SEE WHAT(1h) ˆ ?
*CHI: MOUSE(*N) .
*MOT: t@ag g@ag(nh): \- ˆopr'WHQ WHAT(1h) ˆ \?
*CHI: 0 [%act: lifts panel in book] *mot .
*MOT: ˆopr'WHQ WHAT(1h) ˆ ?
*CHI: < PNT_3[on book] *m > [>] .
*MOT: < ˆopr'WHQ WHAT(1h)(*2) ˆ > [<] ?
*MOT: ˆopr'YNQ \- CAT t@ag(nh) CAT PNT_3(at_cat_in_book) ˆ < CAT > [>] ?
*CHI: < ˆopr'NEG 0 ˆ > [<] [%ges: don't know/not me] .
%ges: open 5s, wrists rotate out
*MOT: ˆopr'WHQ WHAT(1h) ˆ ?
*CHI: <BEAR> [*] [//] BEAR .
%err: BEAR $mvt

1.5   Czech sign language

Czech sign language *is one of the national sign languages*—it is the visual-manual communication system of the Czech deaf (the Deaf). The Act on the Communication Systems of Deaf and Deaf-Blind People No. 155/1998 Coll., as amended by Act No. 384/2008 Coll., defines Czech sign language in Section 4 as follows: "(1) Czech sign language is the basic communication system of those deaf people in the Czech Republic who consider it the main form of their communication. (2) Czech sign language is a natural and wholesome communication system consisting of specific visual-movement means, namely handshapes, hand position and movements, facial expressions, positions of the head and the upper torso. Czech sign language has the basic attributes of a language: it is semiotic, systemic, doubly articulated, productive, independent, and historical, and is lexically and grammatically stable. (3) Czech sign language can be used as a tactile communication system for deafblind people, consisting in tactile perception of its means of expression." *Czech sign language is therefore the natural language of the Czech deaf* (the Deaf). *It is not derived from the main spoken Czech language and has its own vocabulary and grammar.*

Contrastive linguistics—differences between the Czech language and Czech sign language

National sign languages (hence also Czech sign language) differ from the spoken national languages (the Czech language, in this case) in particular in the way they exist and use three-dimensional space. These two fundamental differences between the two language systems grant Czech sign language characteristics which the spoken Czech language does not possess (and vice versa). Differences between the languages are described by contrastive linguistics, which compares the structure, features and specifics of languages.

Czech sign language uses three types of means of expression:

- verbal and non-vocal (movements and positions of hands, face and upper torso);
- non-verbal and non-vocal (gestures, facial expressions, mimetic description);
- non-verbal and vocal (laughter, vocal expressions, speech components).

Phonological, morphological, lexical, and syntactic differences were identified based on findings from research of Czech sign language. (cf. Servusová, 2008)

Phonology

While words in spoken languages (including Czech) are formed by a linear sequence of sounds, signs consist of several smaller phonological units (sign parameters) produced simultaneously. These include namely the sign articulation location (TAB), handshape (DEZ), hand location relative to the body (orientation of the hand and fingers – ORI), hand arrangement (HA) and hand motion (SIG). For more information, see 1.3 Manual factors in sign language.

Sign language phonemes are identified using minimal pairs, which are lexical units (signs) differing in one single phoneme. Sample Czech minimal pairs include ŘEDITEL [DIRECTOR] – SPORT [SPORT] (different TAB), HLAD [HUNGER] – ZÁCHOD [TOILET] (different DEZ), POČASÍ [WEATHER] – ATLETIKA [ATHLETICS] (different SIG), VEČER [EVENING] – OBRAZ [PICTURE] (different ORI1), and MLADÝ [YOUNG] – OBYČEJNÝ [COMMON] (different ORI2). Servusová (2008) points out a theory based on which certain phonemes in sign languages share the same meaning (for example signs articulated at the head are very often linked to mental processes).

Morphology

The system of language units which carry meaning is, in the case of simultaneous sign language, completely different from the system of spoken linear languages. Macurová (2001) argues that the recognition of parts of speech is not important, for example, in Czech sign language, unlike, for example, the directionality of signs, their movement in space, which serves to express grammatical relations. Czech sign language, however, also groups signs under several parts of speech. Major differences in grammar categories are noted with verbs in particular.

Analyses of spontaneous sign expressions of Czech native signers generated the following three groups of verbs based on the way individual signs representing verbs inflect depending on their grammatical meaning (on the specifics of their

inflection) (cf. Macurová, Bímová, 2001; Servusová, 2008):

- **Simple verbs,** which do not change within a sentence, context or to express person (number). They remain in the canonical form and their arguments need to be expressed lexically and sequentially (linearly) with a name proper or pronoun. The grammatical meanings of person (and number) are carried by name proper or pronoun, not by the verb itself. Simple verbs may, however, also change slightly in context, although not in terms of person or number; they can express (mostly non-manually) various circumstances and provide, through incorporation, detailed information about their object. The simple verbs of Czech SL include the majority of verbs articulated on or near the body (e.g. signs indicating TO LOVE, REMEMBER, LIKE, ASK, THANK, SIGN).

- **Compliant verbs** make the maximum use of space, expressing their arguments through a change of form. The grammatical meanings of person (number) is expressed through the incorporation of the place of articulation, the change of direction in the horizontal line and, in some cases, change in the palm orientation. Compliant verbs include signs indicating TO GIVE, TO WATCH, TO CRITICISE, TO VISIT, TO PAY, TO QUESTION, TO TEACH, etc. These verbs used to be referred to as directional verbs.

- **Spatial verbs** do not express the meaning of person (number)—their arguments need to be expressed with a name proper or pronoun. They change in the sentence and context, however, by supplying information about the location of the action or state, about the movement of an object, and often express the means (by incorporating an altered handshape). Spatial verbs make maximum use of the typical features of Czech SL—its existence in three-dimensional space and simultaneity (e.g. signs indicating TO SHAVE, TO WASH ONESELF, TO DRINK). The spatial verb category also includes verbs primarily expressing movement in space (e.g. signs indicating TO FLY, TO GO, TO JUMP). These verbs are sometimes also referred to as locating, movement-locating or classifying verbs.

Irrespective of this division, all verbs in Czech sign language which change significantly with the context (manual and non-manual components) are sometimes referred to as *unstable verbs* in the relevant literature. The *verb aspect* in Czech sign language, in contrast to the Czech spoken language (where it is expressed by grammar only, for example, with prefixes or suffixes), is expressed lexically-grammatically by adding another sign to the basic verb in line with specific grammar rules. In most cases it is the DONE sign, although using the three-dimensional sign space and non-manual components (facial expressions and head movements) is also substantial (cf. Servusová, 2008).

The way time is expressed in Czech SL is highly varied (cf. Macurová, 2003). Comrie (1985, in Macurová, 2003, cf. Černý, 1998) reports that the idea of time is inherent to every culture, and time is expressed in every language. Languages use the two following ways to express time:

- grammatical expression of time, marked in the verb with a specialized verb form (for example past tense in Czech) and
- lexical expression of time using 'time-reference' and time-deixis—means cor-

relating the time referenced with the moment of utterance (for example, "yesterday", "then", "next year").

Czech SL uses the lexical expression of time only, unlike spoken Czech. Instead of being expressed by a change of verb forms, time is therefore conveyed by incorporation and certain temporal signs related to an imaginary timeline. Time in Czech SL, similarly to other (better researched) national sign languages, is expressed along several axial lines, whose referential point is the human body:

- The *front-rear axis* is the basic timeline serving for the articulation of time specifications which place events, actions, situations, states, etc. in the future or the past. The timeline is also used to articulate temporal signs (e.g. the sign of HISTORY, YESTERDAY, TOMORROW).
- The *vertical axis* is used in Czech SL (as in other SLs) to express the time of human growth.
- The *right-left axis* serves for time duration and time shifts.

In addition to the three basic timelines above, Macurová (2003) also describes other, *additional timelines* used for the articulation of other signs referring to time (for example, A LONG TIME and SLOWLY are articulated on the forearm of the non-dominant arm). The timelines are not, however, fundamental for the way time is expressed in Czech SL, as they are a mandatory part (manual component) of specific signs and thus learnt together with other parameters of the sign. As time is expressed lexically in Czech SL, it is usually easy to learn (unlike other grammatical features of Czech SL).

One of the linguistic universalities is the existence of deictic means, which are represented by pronouns, among other parts, both in Czech SL and spoken Czech. They fall, like in Czech, into the following groups (cf. Servusová, 2008):

Definite pronouns:

- *Personal pronouns,* which refer to parties (present as well as absent) to a communication act (I, YOU, WE). As Macurová and Bímová (2001) report, unlike spoken Czech, Czech SL does not distinguish between gender in third person singular and plural pronouns—it is implied by context. In contrast to spoken Czech, other types of personal pronouns anchored in special phonemes (particularly handshapes) are recognized in Czech SL for double, triple and quadrature numbers (YOU TWO, US THREE, THEM FOUR).
- *Possessive pronouns,* which refer to the owner of the given communicated object (MY, YOUR).
- *Demonstrative pronouns,* which point to the given communicated object (THIS, THAT).

Given the visual-manual form and the use of three-dimensional sign space, Czech SL, unlike audio-oral Czech, uses certain pronouns to express the spatial relationships between the subjects and objects the pronouns refer to. Adequate use of these pronouns is an indispensable grammatical feature. Similarly, the plural is more accurately expressed in the pronouns of Czech SL, specifically the inclusive plural (for example WE = I + addressee) and the exclusive plural (for example WE = I + third party) (cf. Macurová, Bímová, 2001).

Indefinite pronouns:

- *Interrogative pronouns,* used to request information from the addressee (WHO, WHAT).
- *Relative pronouns,* whose role is largely played in Czech SL by non-manual means of expression and partly by third-person personal pronouns.
- *Indefinite pronouns proper* refer to an unspecified object (WHAT/WHO + a facial expression of ignorance) (cf. Servusová, 2008).

There are also differences in the structure of numerals. Although research of numerals in Czech SL has not been fully completed, partial studies show that their principle differs vastly from Czech due to the incorporation processes. Drawing on her own research, Šebková (2005) defines *the following four main groups of numerals, of which some are divided into subgroups:*

- cardinal numerals,
- ordinal numerals:
  – indicating dates,
  – indicating sequential order (e.g. in competition, by age),
  – indicating numerical order (e.g. class or year at school)
- multiplicative numerals:
  – indicating how many fold or how many times an event happened,
  – indicating how many times an event happens over a period of time,
  – indicating how many times something is multiplied,
  – indicating how many times an event has happened,
- special numerals:
  – inverted, with the palm facing the front, indicating the last numeral in the year, the school grade, or the city district number;
  – iconic, the signing of which is visually motivated by the shape or position of the number they represent (e.g. jersey number, SQUARE METRE, A SQUARE);
  – specific, whose form or use differs from others (THE ONLY ONE, ONLY TWO, WE THREE, etc.).

Nováková and Šebková (2008) also define the signs of numerals which indicate temporal and financial data. These are incorporated in another sign, which usually represents a noun.

Lexicology

Czech sign language and spoken Czech are equal on the qualitative and quantitative level in terms of the word/sign lexicon (including homonymy, synonymy, polysemy, antonymy), as both Czech SL and spoken Czech are able to express any meaning. Czech SL has, however, a different system for categorizing vocabulary (it uses different signs for some of the polysemal Czech expressions and lacks selected signs for hyperonyms, for example). It also differs in the creation of new signs, in phraseology, and above all, in the substantial use of facial expressions on the lexical level.

Signed languages, similarly to spoken languages, adopt word-stock from other linguistic systems. Although multiple language users do not agree with the trend,

it is natural and cannot be removed. While spoken Czech usually adopts foreign words undistorted, including their pronunciation (e.g. INTERNET or AUTO) or adopts them to Czech speech (e.g. DISPLEJ, RAKETA), signed languages convert them to another existential mode (audio-oral mode is converted to visual-motor mode).

The **signs of Czech SL** (the smallest lexical units carrying meaning) are divided into the following three groups based on their nature and origin:

- *Demonstrative (deictic) signs*—they are the simplest and most comprehensible signs of Czech SL. They point to objects (representing their meaning) in the immediate vicinity of the speaker (e.g. signs indicating EYE, EAR, TEETH).
- *Imitation (iconic) signs*—their course and nature resemble the typical characteristic of an object or phenomenon they represent (they are visually motivated). Their meaning is therefore easy to comprehend. Iconic signs enable communication about objects which are physically not near the speaker (e.g. signs indicating a car, HOUSE, TO SLEEP, TO EAT). Iconic signs are divided into the following groups:
  - *transparent signs,* whose visual motivation is clear, and the sign is comprehensible even for non-signers (e.g. TO DRINK, TO RUN, TO BRUSH ONE'S TEETH);
  - *translucent signs,* which are visually motivated, and yet meaning is comprehensible only to signers (e.g. the signs for YEAR, SOCK, OCTOBER). The distinction between transparent and translucent signs is blurred, mainly because it is impossible to objectively quantify the degree of visual motivation of a particular sign.
- *Arbitrary signs*—abstract, agreed signs, which evolved for a long time. They refer to abstract qualities and events and need to be taught to make sense (e.g. the signs for TUESDAY, SON, ILL).

Traditional vocabulary forms of signs may, however, change significantly due to incorporation, context, and current communication situation in everyday communication in Czech SL. The typically iconic sign of CHAIR, for example, may be replaced by simply pointing to a chair if the chair the speaker mentions is near said speaker. In this case, a deaf signer spontaneously uses a deictic sign to denote a particular chair.

A special component of the lexical aspect of sign languages is *specific signs,* which are a *typical part of the intra-cultural communication of the Deaf,* namely of native signers (cf. Vysuček, 2008). Specific signs, also referred to as *multichannel signs* or *Deaf Signs* (cf. Lawson, 1983; Engberg-Pedersen, Hansen, 1986), do not have any exact counterparts in spoken languages (such as Czech). Their semantic meaning needs to be described using multiple words, and the meaning often varies from the perspective of the spoken main language—exact meaning depends primarily on context (hence specific signs). They often carry a positive or negative emotion, express a mental state, the attitude of the speaker to someone or something or the internal evaluation of something.

Syntax

Czech sign language is simultaneous, and the arrangement of words (signs) in a sentence in Czech sign language differs considerably from spoken Czech in the use of incorporation, classifiers, word (sign) order, active use of three-dimensional space, non-manual means in syntax and the communication function of the utterance.

Formation of interrogative sentences is considerably different on the syntactic level in the Czech language and in Czech SL. Questions in Czech SL, similarly to spoken Czech, fall into *yes/no questions* and *WH-questions*. In addition to lexical means, *the non-manual component of a sign and the location of interrogative expressions* (pronouns, adverbs) play a key role in Czech SL, unlike in spoken Czech. Facial expressions (raised eyebrows and intense visual contact in particular) distinguish statements from interrogative sentences. In addition to being a distinguisher, the non-manual feature serves pragmatic (alerting that a response is expected), regulatory (signalling the beginning and end of a conversation), and reference (identifying and locating the speaker) purposes.

The basic lexical means contained in questions in Czech SL are *the interrogative expressions (pronouns and adverbs).* These include, according to Hronová (2002), WHO, WHAT, WHERE, WHEN, HOW MUCH, HOW MANY, AND WHICH. These interrogative expressions do not inflect in Czech SL, unlike in spoken Czech. Their form remains unchanged, in contrast to, for example, compliant verbs in Czech SL which autonomously express the grammatical person and number of the subject and object, without having to use a name proper or pronoun to indicate meaning. Analysis of the spontaneous communication of Czech deaf people revealed that some of the interrogative expressions which are autonomous in spoken Czech are replaced in Czech SL by a single umbrella interrogative sign and vice versa (e.g. the sign for WHO = the first, second, third, and sixth case of Czech 'who'; the Czech "which" is signed as WHO, WHERE, WHAT in Czech SL depending on context). Students of Czech SL need to be aware that questions in Czech SL do not follow the word order and grammatical structures of the Czech language, nor are questions created by replacing Czech words with sign language counterparts.

An essential specific property of Czech SL, which is common in everyday Czech SL communication is incorporation, which takes advantage of the simultaneity and existence of sign language in three-dimensional space. Incorporation refers to the penetration of other qualities into the vocabulary forms of Czech SL signs and to their contextual and content modification using manual and non-manual components (several different signs are incorporated into one another). Incorporation and the non-manual components of signs help express the grammatical component of utterance. It is an inherent and integral part of Czech SL.

Several types of incorporation are used in Czech SL, for example:

- time incorporation (for example "last week" can be expressed in Czech SL with a single integrated sign composed of modified signs for LAST and WEEK);
- *quantity incorporation* (e.g. 3 HOURS, 2 MONTHS, 4 YEARS);
- *manner incorporation* (e.g. RIDE FAST/SLOWLY);

- *incorporation of the subject into the verb* (e.g. DRINK FROM THE BOT-TLE/TAP/CUP, SHAVE WITH AN ELECTRIC/MANUAL RAZOR);
- *direction incorporation* (e.g. RETURN TO ME/SOMEONE, ASK ME/SOME-ONE);
- *negation incorporation* (for example, by articulating the sign I HAVE while shaking the head from side to side to convey "I don't have").

Incorporation can also be classified based on the parts of speech concerned (cf. Servusová, 2008):

- noun + numeral,
- pronoun + numeral,
- noun + adjective,
- verb + adverb,
- verb + noun,
- verb + pronoun.

Hearing people who are starting to learn Czech SL often struggle to understand deaf Czech signers when they use incorporation as the incorporated signs differ substantially in many cases from the vocabulary form which the hearing are familiar with. On the other hand, the failure to use incorporation by the hearing makes signing difficult to understand for the deaf as the resulting language is Czech SL, or spoken Czech supported with signs.

Another difference between Czech SL and spoken Czech is *the use of classifiers*. Classifiers are morphemes associated with other signs and representing a certain meaning (cf. Karlík, Nekula, Pleskalová et al., 2002; Černý, 1998). They are referred to as *representative signs*. In spoken languages, classifiers most often represent shape, material, consistency, size, layout, quantity and location. In Czech SL, classifiers are used mainly to specify information about the location of objects, their size, shape and also to facilitate spatial orientation. They are most often associated with verbs and nouns. Classifiers also serve as vehicles for the introduction of new signs in Czech SL, enriching the language's lexicology.

Classifiers are grouped into *four elementary types* (Motejzíková, as cited in Servusová, 2008):

- *whole entity classifiers*—representing objects with certain identical features;
- *limb classifiers*—they describe the motion of human or animal limbs based on what is typical for them;
- *handling classifiers*—they represent the size and shape of an object by the way it is held and moved;
- *shape and dimension classifiers/specificators*—they give a detailed characterization of the appearance of objects.

The possibility of classifiers and incorporation grants Czech SL with another specific property—*a mimetic description (space visualization)*, which enables a detailed, simple, and rapid description of the features of a given object. The visual-motor three-dimensional nature of sign language allows for active use of the available signing space. Holubová (2006) argues that space and its use is necessary in sign languages due to the visual-motor nature of sign language itself and because it participates in grammar—sign morphology, syntax and semantics.

Sign languages distinguish between a topographic space and a syntactic space. The *topographic use of space* corresponds to the actual distribution of objects (things, people, etc.) mentioned in the communication, including their spatial context (such as mutual position, distance, and height) displayed on a certain scale. The *organization of syntactic space* is, however, arbitrary, and the meaning of spatial contexts is abstract (related to a given reference point, which represents a certain semantic sign) (cf. Engberg-Pedersen, 1986).

Variation of the Czech sign language

Several varieties of Czech sign language are used, as in the case of spoken Czech. The national spoken Czech language has, for example, a formal variety (standard Czech), used mainly in official communication and in written form, and a conversational variety (colloquial Czech), which people use in personal communication. Speakers of any language (including Czech SL) spontaneously alternate between different language codes and varieties, and depending on the particular situation, diglossia is often observed. *Diglossia* maintains social intensity and solidarity within a community (cf. Homoláč, 1998).

The Encyclopaedic Dictionary of Czech (cf. Karlík, Nekula, Pleskalová et al., 2002, p. 113) defines diglossia as "a stable linguistic situation in which all the speakers of the same linguistic community use two variants of the language—one tends to be socially considered higher and is not used in everyday communication, while the other has a lower standing and is the mother tongue of the speakers." Surveys of American Sign Language (ASL) reveal the following findings: "...diglossia determines that English is an adequate language for communication with hearing people (outside people), while ASL is suitable for communication with deaf people. Diglossia serves as a buffer between the hearing community and the deaf community" (cf. Homoláč, 1998, p. 7).

The variability of Czech sign language is natural, developing over time and manifested primarily in its lexical component (stock of signs).

The most common causes of the varieties include:
- natural development of signs;
- influence of Czech SL grammar (incorporation etc.);
- regional aspect (similar to regional dialects in the spoken Czech language, different areas of the Czech Republic use different signs of the same meaning);
- social aspect (educated deaf people highly competent in Czech incorporate elements of Czech into their signing);
- generational and age aspect (younger Czech signers use different signs in some situations compared to older signers; baby sign language is significantly simpler and more transparent compared to the sign language of deaf adults).

General information about the features of sign languages and about Czech sign language in particular suggests that Czech SL is an integral language system in compliance with the definition set out in Act on the Communication Systems of Deaf and Deaf-Blind People No. 155/1998 Coll., as amended by Act No. 384/2008 Coll. Czech sign language is also a basic means of communication be-

tween teachers and children with severe hearing loss, and is an integral part of the education systems of simultaneous, total, and bilingual communication (cf. Gregory et al., 2001; Hrubý, 1999; Jabůrek, 2004).

Research into Czech sign language

Research on national sign languages has grown considerably over the past decades in many countries around the world. The Czech Republic launched *research into Czech sign language* as late as 1993, and the majority of it is carried out by the group around the linguist Alena Macurová. The original goal of the research was tied to practice, as schools for children and youth with hearing impairment had begun to point out the inconsistency of signs of Czech sign language and called for their unification and a general definition of which signs were "correct" and which not. The call concerning the lexical component of Czech sign language cannot be fulfilled until Czech sign language has been described in full, including all its components. The research led by Macurová is aimed at collecting all the data and provides *a thorough description of Czech sign language*.

According to Macurová (in Červenková, 1997), Czech SL research is mainly focused on the following three overlapping areas:

- the use and distribution of Czech SL, its social status and attitudes towards Czech SL;
- the structure of Czech SL;
- Czech SL and thinking.

The launch of the research on Czech sign language rested on three elementary fundamentals:

- the first is the assumption that Czech SL is a complete human language with a structure comparable to spoken languages (the Czech language, for example);
- another fundamental assumption is that Czech SL and spoken Czech have equal functions, in addition to a comparable structure—each language is adapted to fulfil the tasks it is assigned;
- the final elementary fundamental position states that Czech SL language and spoken Czech are fully independent from one another and unrelated.

Macurová (in Červenková, 1997) also presents three levels on which the linguistic study of the structure of Czech SL is based:

- The first is the basic differentiation and description of the fundamental language units without a meaning of their own but which contribute to the differentiation of the language (phonemes). It is therefore necessary to identify and describe all meaning-distinguishing handshapes, the different sign locations, the types of movement, etc. and define what creates meaning and what is a mere variant. The identification and excerption of Czech SL phonemes are conditioned by the scientific notation of the canonical form of the signs—notation of Czech sign language (see below).
- Another level is the study of the lexical aspect of Czech SL. The signs are studied in terms of their regional and generational differences; research also focuses on the semantic relations between signs, the methods of their derivation, the development of new signs, etc.

- The third area of research consists of Czech SL grammar, the relations between signs and the ways they are combined into higher units (sentences).

The primary objective of the research is a comprehensive description of Czech sign language (the lexical and grammatical aspects). The methods of research and linguistic description of Czech SL are vastly different from methods used in the research of spoken languages due to the different mode of existence and specific features of sign language. Major difficulties in the research of Czech sign language become apparent, according to Macurová and Bímová (2001), already during the collection of linguistic data. The best option of the researcher being a native speaker of the language under study is not available in the Czech Republic due to the shortage of linguistically trained instructors of sign language. Elicitation method was therefore used to describe Czech SL, although it is problematic and its credibility questionable.

The problem and pitfall of the elicitation method is, in particular, the fact that the language data are obtained from native users through asking questions in the language of the researcher. It is assumed that informants (native speakers of the language under study) have a good knowledge of the language of the researcher, understand its structures and are able to find counterparts in their language. Czech deaf people are, however, not very competent in spoken Czech. Another problem with the elicitation method is how to ensure a true representative sample of the autonomous language under study. The language may reflect the language of the researcher (language interference) as "users of low-status languages (including sign languages) 'upgrade' their responses in an 'elicitation situation' towards more prestigious language variations, and in the case of sign languages toward varieties closer to the majority language" (cf. Macurová, Bímová, 2001). The danger is high in the case of the research of Czech sign language.

While other research methods remain unsuitable for the study of Czech sign language, the first Czech deaf people trained in linguistics have begun participating in research in recent years. Macurová designed a system inspired by linguistic studies of American Sign Language to describe the lexical units of Czech SL (signs) (cf. Macurová, 1996; Okrouhlíková, 2008).

Comprehensive research of Czech SL is a long-term project. It is vital, however, to the acquisition of high-level communication skills, as a full application of total and bilingual communication systems in the education of children with hearing impairment requires excellent instruction material used for teaching the sign-stock and, particularly, the grammar of Czech SL. Such textbooks are not available on the Czech market yet.

The status of Czech sign language

Sign language can be implemented in the everyday life of the deaf and approaches using sign language (bilingual communication, total communication and simultaneous communication) can be introduced in the education of the deaf only if *sign language is recognized as an independent language* equal to spoken languages. *The status of Czech sign language* is, however, low due to being officially suppressed and prohibited for many years. Improving its position is a challenge for

years to come. Although sign languages are a natural means of communication for people with hearing impairment, they have always had to "fight" for recognition in the majority hearing society. This was mainly due to the fact that the hearing were unwilling to accept that sign languages were different and grant them the status of independent language systems.

The current status of Czech SL and its development is the result of many years of suffering from *official repression and prohibition*. Czech SL was purposefully degraded in the eyes of the public and the deaf. The persecutions prevented it from developing naturally; nor was it ever researched linguistically (in contrast to for example ASL, British Sign Language, and Scandinavian sign languages). This extreme approach has led to the fact that part of the deaf community themselves do not consider sign language to be a language (despite communicating in it happily).

The importance of sign language was subsequently confirmed on the international level. Expert collaboration between ten countries prompted *UNESCO to issue the following resolution in 1984*:

1. the deaf child should have access to both spoken language and sign language,
2. sign language should be recognized as a legitimate linguistic system and enjoy the same status as other linguistic systems.

Under continuing pressure on the UN from the World Federation of the Deaf to take sign language under its auspices as part of its language minority policy, and prevent its possible suppression by oralism, *the European Parliament officially* approved *a sign language resolution* for the twelve countries of the subsequent European Union *in 1988*. It recommends:

1. recognizing the sign language used in every Member State;
2. seeking ways to remove obstacles sign language faces;
3. encouraging Member States to introduce measures to establish, support, and develop a) research into sign language, b) sign language instruction, c) sign language interpreting, etc. (cf. Jabůrek, 1998)

   Influenced by new scholarly findings published abroad, the Czech Republic began implementing the recommendations and resolutions above in the early 1990s. The Czech Federal Assembly added Paragraph (2) to the then School Act No. 171/1990 Coll., Section 3, on 3 May 1990, which read: "Deaf and blind people have the right to education in their language using sign language or Braille" (cf. Hrubý, 1999, p. 327).

One of the most significant initiatives improving the status of Czech sign language was the *adoption of the Sign Language Act* No. 155/1998 Coll. in 1998, which recognizes Czech sign language as equal to other languages, such as Czech, German and English. Deaf citizens have, similarly to hearing citizens, the right to obtain information in their language. The law grants people with hearing impairment the right to communicate with hearing people in sign language. Sign language fully replaces spoken language in the communication of deaf people, and is as indispensable to people with hearing impairment as spoken language is to hearing people. The law states that deaf people are entitled to communication support when visiting a doctor, attending court, at a public meeting, etc. A special

provision of the Ministry of Education, Youth and Sports of the Czech Republic set conditions for free communication support for deaf secondary school and university students.

The Sign Language Act was amended and renamed ten years later as the *Act on the Communication Systems of Deaf and Deaf-Blind People* No. 155/1998 Coll., as amended by Amendment No. 384/2008 Coll. The amended version of the Act mentions the tactile version of Czech SL for the deaf-blind (and other specific communication systems for people with deafblindness) and removes the term "sign language" from the entire wording of the Act, including related laws.

Research into Czech sign language has opened up broad possibilities. *The status of Czech sign language has been rising* and its current linguistic research and increasing popularity among the lay and professional public point to further growth and development. The future of Czech SL depends on all the stakeholders who are in direct contact with Czech SL (parents, experts, teachers of children with hearing impairment, and, in particular, the deaf community). A positive trend is that today's undergraduate deaf education programmes provide theoretical and practical training and require the students to learn Czech SL. Research into Czech sign language has opened up broad possibilities. Introducing sign language into educational programmes, raising public awareness, and promoting its wider social use also depends on its further development and improvement. Its status will grow with the creation and documentation of the stock of signs (sign language dictionaries and video recordings) and with further research into the language system. Equally important is popularization and instruction of sign language.

1.6   The study and lexicography of sign languages

Comprehensive and effective communication, and hence the harmonious development of deaf children, is only possible if the parents and educators of children with hearing impairment are able to adequately communicate in the relevant sign language. Given the specifics, the study of sign language differs considerably from the study of other foreign languages. Macurová (2001) explains that truly successful acquisition of sign language is possible if the student accepts its otherness and understands that "other" does not mean "worse". While learners studying a typical foreign language (e.g. French or German) have no doubt about the independence and distinctiveness of the language (own lexicon, grammar, history, etc.), in the case of a national sign language the subject is underestimated and undermined, notably in relation to the national spoken language.

The fundamental contrast between spoken foreign languages and sign language is the different mode of existence. *Sign language requires learners to use means of expression which they have not yet had to use (or to a small extent) in their communication in spoken language, and which consequently do not feel natural.* The means include the exact and sometimes complicated positions of palms and fingers, movements of hands, head and whole body, use of three-dimensional space and facial expressions. Hearing users of a spoken language (which is linear) also struggle with the simultaneity of sign language.

Sign language is usually a secondary language (their primary, mother tongue is the national spoken language) for hearing learners, and as such it is considered a "second" language. The authors of the ASL teacher handbook (in Jabůrek, 2000) define the following basic principles of the instruction and study of a secondary language:

- the aim is the ability to communicate in the target language;
- the most fundamental factor in learning any language is appropriate contact with the language;
- the student needs to be inspired, motivated, and have an opportunity to communicate in the target language;
- dialogues, practising, communicative activities, and games are the basis for meaningful contact with the target language;
- language instruction, specifically with adults, is partly a conscious process, not a mere imitation of the instructor. Students should be aware at all times of what they communicate;
- mistakes are made during instruction, and they are an important and integral part of the process.

Macurová (2001) argues that, as with any foreign language, instruction of sign language needs to focus on the stock of terms (signs) of the language (learning signs as headwords) and on its grammar (learning how the canonical forms of signs change in sentences and how sentences are built in sign language). *Learning the signs* (their manual component, to be more precise) *is the easiest part of studying a sign language. It is, however, far from the only aspect students need to learn to master the sign language* (as it is impossible to learn English, for example, by studying a Czech–English dictionary).

Focusing on the lexical component of sign language and underestimating its grammar results in the acquisition of communication competence in the signed national language (Czech SL in this case), instead of the national sign language. Grammar is conveyed in sign languages by facial expressions and movements of the head and upper torso. "The movements and positions of these body parts do not play a current role in sign language, something which may, but does not need to accompany what the hands express. On the contrary: non-manual components are a 'mandatory' part of sign language and have grammatical validity—they carry grammatical meanings, similarly to endings in spoken Czech" (cf. Macurová, 2001, p. 74).

The acquisition of communication competence in sign language is conditioned by *systematic theoretical and practical study*. Understanding the specificity, features, and theoretical structures of the sign language facilitates and accelerates the practical training and acquisition of the language. Practical skills in sign language, in turn, improve insight into its theoretical background. The two components—theoretical knowledge and practical skills—need to be developed simultaneously and on a complementary basis. Another prerequisite for mastering sign language is *knowledge of the specifics of the culture and communication standards* of its natural users, namely the deaf (the Deaf).

Hearing people have three options in terms of acquiring communication com-

petence in sign language: *self-study, a sign language course* and *spontaneous communication* with a deaf signer. Each method has its advantages and disadvantages, and the best results are achieved by combining them systematically. The fastest way to acquire communication competence in sign language is to take engaging language classes. High-quality language courses are taught by sign language instructors who are sufficiently competent in sign language and have adequate teaching skills. The best instructors in terms of language competence are deaf signers. Deaf instructors are, however, often inadequately trained in sign language theory and lack the necessary teaching competence. Hearing, pedagogically competent, and knowledgeable instructors, on the other hand, often lack sufficient communication competence in sign language, teaching signed language instead of sign language. Instructors with a full linguistic and teaching competence are unfortunately rare.

Potentially deficient competence in sign language instructors could be at least partially eliminated if the instructors had high-quality methodical material available, which would offer information and suggestions they could apply in class. Students should be able, however, to use high-quality instructional material. *Effective study of sign language and adequate communication competence thus require a comprehensive and dynamic approach.* Self-study supported by instructional material should be conceived as preparation for direct communication and revision of what the student has learned. Sign language courses led by a competent instructor help the student quickly acquire a certain communication competence, which can only become advanced following direct and repeated communication with the deaf.

The *lexicography of sign language* is responsible for compiling sign language dictionaries, which are a useful tool for studying and later communicating in sign language. The first sign language dictionaries began to appear at the end of the eighteenth century, drawing on the work of Abbe de l'Épée, who incorporated sign language in the education of people with hearing impairment. Signs were primarily described with words, later supplemented by pictures of the sign. The first dictionary of sign language compiled in line with the linguistic lexicographic principles was published in the USA in 1965 by William Stokoe and his team (Okrouhlíková in Slánská-Bímová, Okrouhlíková, 2008). The first publications and dictionaries of Czech SL and Czech SL (referred to as "sign language" at that time) appeared in the Czech Republic (Czechoslovakia) in the 1980s. The total production published to date includes about twenty books and serial publications, five video and DVD programmes, fifty multimedia CD-ROMs and DVD-ROMs, and four online dictionaries.

Although the number of dictionaries and materials designed for the study of the signs of Czech SL and signed Czech is high, none are focused on teaching Czech SL grammar. Face-to-face instruction and communication with signers is the only option available to learn the grammar of Czech SL.

Printed textbooks and vocabularies (dictionaries) remain popular among students of foreign languages. The main reason is that they are easily available and cost-effective. The requirements and demands on sign language dictionaries vary

considerably from common foreign language dictionaries due to the specific features of sign language. Given the current technology, the printed form is outdated and inadequate in the case of sign language dictionaries. Books, by their nature, allow for only a static description of information, while sign language is a visual-motor communication system. As it uses three-dimensional space as well as non-manual components, it is difficult to capture the sign implementation accurately and precisely if we are limited by the immobility of the image. All manual components of a sign, including the movement of the hand (or both hands), can only be captured in books with a verbal description of the implementation of the sign and a still photograph (or picture), complemented with arrows or other auxiliary marks. The non-manual components of signs are nearly impossible to capture, apart from a not particularly exact verbal description.

Sign language educational videos began to emerge in the early 1990s. The videotape, as a carrier of signs of the relevant sign-stock, is able to accurately capture the implementation of the interpreted sign, including its non-manual components, in comparison with the static images presented in books. Unlike a book, it is able to illustrate how isolated signs are combined into coherent messages, incorporation in sign language, work with space, classifiers, mimetic description, and other specific features of sign language. Coherent narrations (dialogues) serve to illustrate the grammar structures of sign language using examples. A high-quality video player allows for slow motion playback as well as frame advance and freeze frame, which is useful for detailed study of the utterance or sign.

The limitations of VHS tutorials are, apart from the requirement to use costly technical equipment, the poor search options, specifically when searching for a particular expression, and the gradual wear and tear of the tape and video recording. VHS tutorials have been replaced by DVDs, but these are not very popular as modern computer technology seems to be much more advantageous.

The multimedia nature of electronic sign language dictionaries enables the combination of audio, text, and visual (images or videos) communication, which the previous forms of dictionaries were unable to offer. *The interpretation of signs with digitized film clips is dynamic, faithfully capturing the implementation of a sign*, which can be played back at any speed or frame by frame. The high operational speed of current modern computers also enables rapid searching in the dictionary, organization of signs into thematic units or lessons, and searching for specific signs according to a variety of criteria (Langer, 2007, 2008). Similarly to videotapes, multimedia are able to present both isolated signs of Czech SL and narrations and dialogues. It is the best instructional material designed specifically for self-study and homework (cf. Langer, 2006).

In addition to rapid orientation in an electronic dictionary, easy retrieval of entries, and the ability to play the sign recording clip at any speed, multimedia instructional material offers many other features. All signs can be viewed from several angles (the front and the side perspective by default, possibly also a detail) so that all the phases of the sign are clearly visible. The meaning of the signs or their Czech equivalents can be clarified using brief definitions, pictures, spoken commentary or another short video explaining the meaning in sign language.

Practising and verifying the signs the student has learned is useful from the didactic perspective. *The experience of users of multimedia dictionaries and sign language instructional material suggests that electronic products are highly relevant and beneficial for the study of Czech sign language.* User feedback indicates that multimedia products facilitate intensive self-study and thus allow for using face-to-face instruction for practice and conversation.

References to law acts and internet resources

Law acts:
1. Zákon o komunikačních systémech neslyšících a hluchoslepých osob [Act on the Communication Systems of Deaf and Deaf-Blind People]. Sbírka zákonů, č. 155/1998 ve znění novely č. 384/2008 Sb.
2. Zákon o sociálních službách [Social Services Act]. Sbírka zákonů, č. 108/2006.
3. Zákon o znakové řeči [Sign language act]. Sbírka zákonů, č. 155/1998.

Internet resources:
1. www.amazon.co.uk
2. www.cktzj.com
3. www.cztn.cz
4. www.derwen.ac.uk
2. http://inthelandofinventedlanguages.com
2. www.jacobbailey.com
2. www.signwriting.org

Peter Angermann, *Homunculus*, 1989, watercolour

2 QUANTITATIVE LINGUISTICS AND BEYOND

2.1 General information

> When you can measure what you are speaking about,
> and express it in numbers, you know something about it.
> When you cannot measure it, when you cannot express it in numbers,
> your knowledge is of a meager and unsatisfactory kind;
> It may be the beginning of knowledge,
> But you have scarcely, in your thoughts, advanced to the stage of science.
>
> Lord Kelvin

To be able to express what we measure in the realm of linguistics in numbers, we use services of mathematical linguistics which attempts to describe and model the language with mathematical, mostly quantitative methods. The methods to be used which are at hand are either symbolic (algebraic, formal), statistical or generally quantitative, or a combination of the mentioned. A very significant function with which mathematics can contribute to the development of linguistics is precise setting of hypotheses once qualitatively felt and enunciated in a quantitative way which is the only way to have the chance to verify them or falsify.

The VIIIth International Linguistic Congress in Oslo in the year of 1957 is regarded as the official origin of mathematical linguistics. Mathematical linguistics as such can be seen as an opposite to qualitative linguistics, one might say an opposite to "traditional linguistics". We would like to prove here that traditional might not always mean qualitative by an insight into the history of quantitative linguistics. And we would also like to mitigate the tension which is sometimes felt between qualitative and quantitative linguistics.

Currently, mathematical linguistics is understood as four independently developing scientific branches, i.e. quantitative linguistics (having sometimes been called also as statistical linguistics for its ample involvement of statistical instruments; we, nevertheless, would like to stress that "statistical" need not always mean "quantitative", yet it is a subset, cf. further in the text), algebraic and machine linguistics. Corpus linguistics is the fourth one. The most swiftly developing branches are machine linguistics due to worldwide computer utilization, and quantitative linguistics whose instruments are, by the way, employed in this research. Thus, we will, in a few words, first discuss the other branches.

Algebraic linguistics (cf. e.g. Sgall et al., 1973), does not use statistical methods for linguistic research. Its origins can be found in the second half of 19th century, when formal logic developed dramatically and it was accompanied with abstract and non-quantitative procedures. The name of this discipline was proposed by Y. Bar-Hillel in the second half of the 1950s. Algebraic linguistics occupies with the analysis of arranged chains of language units, above all on the syntactic language level. To name at least a few iconic works—among the most known and most important there are the generative grammar by Noam Chomsky (cf. Chomsky, 2002), functional generative description by P. Sgall, application generative model of the language by S. K. Šaumjan, recognoscative and categorical grammar, analytical language model, dependency grammar and others.

In the second half of the 20th century, the development of information technologies started to rocket, which was caused directly by accelerated growth of information volume which called for being processed or at least stored. Machine (sometimes called computer or computational) linguistics links and applies practically theoretical quantitative and algebraic linguistics. Working with computers and searching for information from worldwide brought the quest for text interpretation. The necessity to find the way to interpret texts quickly using computers proved in a way to be a big challenge. The problem appeared above all in the semantic component of natural languages which has resisted any particular description most and therefore resists also to the transfer to any other natural language via computer. However, machine linguistics assists also in other activities, e.g. the spectral analysis of speech, processing frequency lists etc. When processing these, we rely above all on computer accuracy, capability of computers to process huge amounts of data and on computer operation speed. In connection with the use of text analysis and data in general in the humanities, the subject of Digital Humanities is now being discussed. This discipline is not methodically defined, it takes the tools for material processing from the disciplines discussed above, the common goal of this discipline is to discover new knowledge in humanities research through the processing of large amounts of data or more complex data analysis.

Quantitative linguistics focus is defined by Marie Těšitelová in the following way: "Quantitative linguistics is a subset of mathematical linguistics which quantifies (finds out quantitative data) phenomena of various language levels and models their relations realized in the sentence, in the text so that we are able to discover their causal mechanism, their operation, their formal aspect, but also the semiotic one. Since statistics has mostly been used when applying quantitative methods in linguistics, it is sometimes denominated as statistical linguistics. It is, however, a denomination narrower than the one of quantitative linguistics. Therefore, it is not possible to exchange one for the other in all cases" (cf. Těšitelová, 1987, p. 8–9).

This branch of mathematical linguistics, recognized for applying quantitative methods, is deep-rooted already in the work "Janua linguarum reserata" by Jan Amos Komenský (cf. Komenský, 1959), where he showed how it is possible to use the knowledge of word frequencies for foreign language teaching. Among others who drew attention to employing quantitative methods in linguistics in the

19th century, there definitely was also German mathematician and one of the most important members of the Russian mathematical school V. J. Bunjakovskij. At the end of the same century neogrammarian Herman Paul enunciated that the language is a statistical average of language manifestations of all its users; then, he uses the term of an invariant sound in phonetics. A quantitative point of view can be found in an interesting work by Martin Hattala – Initial consonant clusters in Czechoslovakia (1870). Hattala combines phonotactic and quantitative aspects and foreshadows interest in frequency in word-formation issues. Frequency as an explanatory factor of language development appear rarely until the sixties of the twentieth century—e.g. Igor Němec (1971) explains the development of synonymous word-forming means based on the frequency of members of phonological opposition in suffixation a hundred years later in a very innovative way and in connection with the theory of information. As far as language development and diachronic linguistics are concerned, glottochronology has become the most influential application of quantitative methods in this field. This lexicostatistic approach uses words that are resistant to borrowing to express the genetic kinship of languages. This is done by a quantitative analysis of a list of such words (loanwords) from different languages. The Swadesh list (see Swadesh, 1955) is used for this purpose, its modifications are currently being used (see Leipzig–Jakarta list, Haspelmath, Tadmor, 2009).

In the 80s of the same century, mathematician August Seydler used probability instruments to decide on authenticity of so called "Rukopis královedvorský a zelenohorský" (cf. Seydler, 1886). Almost a hundred years later, in the year of 1962, Americal statisticians Frederick Mosteller and David Wallace attempted to find in the language of texts in Federalist papers structures assisting to determine the author by means of the power of Bayesian inference and the computational power, extraordinary for that time. Quantitative methods of text attribution are represented by Vašák, 1980—these are methods of text attribution based on text metrics such as type token ratio, average sentence length, etc. Holmes, Robertson, Paez, 2001 talks about traditional and non-traditional determination of authorship—authorship attribution on the basis of quantitative analyses is considered as innovative, in this period mainly associated with the use of function words and their distribution in texts, using the methods of Machine Learning or natural language processing (see Rudman, 1998; Burrows, 2003; Binongo, 2003; Baayen et al., 2002; Juola, Baayen, 2003; Diederich et al., 2003; Zhao, Zobel, 2005). A quantitative analysis of the text (e.g. Yule, 1938, 1944) heralded this rapid development of authorship attribution in relation to the distribution of linguistic phenomena in the text—e.g. the Federalist letters analyses, which is a famous case in this field of study (cf. Juola, 2006). For the authorship attribution the distribution of linguistic phenomena in the text reflects the unconscious authoring means of text creation. In connection with the use of quantitative methods for the analysis of a literary work or text in general, it is referred to as stylometry.

In the 1870s, William D. Whitney studied as one of the first linguists frequencies of (English) vowels. Incidence frequency is one of the most important step-

ping stones of quantitative linguistics important not only for linguistics itself, but also for e.g. methodology of language teaching (how to pick up the most important vocabulary for language students on different levels), until recently also for printers, stenographers (the busiest words match to the simplest symbols), for designing and producing of various games and brain teasers etc. Also Samuel Morse used the knowledge of letter frequencies when designing his alphabet; he chose the simplest sign for the most frequent letter etc. The first frequency dictionary appeared at the very end of the 19th century. It was gathered by German stenographer F. W. Käding (cf. Käding, 1897). The development of phonetics at the turn of the twentieth century enjoyed a statistical perspective very much—it is related to the experimental nature of the discipline. In this context, statistical aspects have also been reflected in the formulation of basic structural concepts—although structuralism is largely ignoring the quantitative point of view in linguistics. An example is the concept of flexible stability (or dynamic equilibrium) which describes relationship of language system (langue) and linguistic use (parole) by V. Mathesius (1970), where the abstract system of linguistic values in the sense of the phonological system is derived based on the different form of units in the text.

At the beginning of the 20th century, Russian mathematician Andrej A. Markov contributed significantly to the development of quantitative linguistics with his analysis of "Eugene Onegin" (cf. Markov, 1913), where he concluded that in every utterance part, it may be anticipated with a certain level of probability which language units would follow. The information volume can, therefore, be measured—this significant theory is called the Markov process. Speech is then according to Markov a process in which the already produced language units are followed by others by the relative frequency which is mandatory for the given language. The relation holds for unit frequencies (including also gaps), probabilities that a unit is followed by another and the fact that different units bear different information rate. This knowledge is closely linked to the so called theory of information. It says that the highest information rate is borne by the unit which can be estimated hardest in advance. The probability of language unit incidence is directly proportional to the frequency of the given unit in the language.

This research was contributed in the 1920s and 1930s by American linguist of German origin, Harvard university professor George Kingsley Zipf. He studied the relative vowel frequency and found out that vowels and their classes in different texts have the same frequency. In all languages, the number of voiceless vowels is approximately two times higher than the number of voiced vowels. And the more difficult it is to articulate vowels, the lower frequency they have, which is connected to the principle of economy, of least effort in the language, i.e. the theory that one single primary principle in any human action, including verbal communication, is the expenditure of the least amount of effort to accomplish a task; (cf. e.g. Zipf, 1949; Benešová, Faltýnek, Hadwiger, 2014; Piantadosi, 2014).

The other half of the 20th century brings a huge development of quantitative linguistics which is to a high extent caused by the simultaneous development of information technologies enabling processing of huge amounts of data automatically. For nice surveys in this field, see e.g. Köhler, Altmann, Piotrowski (2005);

Köhler, Hoffmann (1995); Tuldava (1995). Czech linguist Luděk Hřebíček (for more details see below) writes that "the real linguistics of the other half of the 20th century is Altmannian linguistics" highlighting the contribution of German quantitative linguist, born in Slovakia, Gabriel Altmann (cf. Hřebíček, 2008). Professor Altmann always highlighted precise formulation of scientific laws and testable hypotheses, i.e. hypotheses which can be verified or falsified at best by means of statistical methods. Altmann also made the formulation of several linguistic laws originally formulated more or less in a qualitative way precise so that they could be tested in a quantitative way, e.g. the Menzerath–Altmann law, cf. further in the text. Gabriel Altmann is, by no means, one of the founders of quantitative linguistics.

Not only in the context of Czech linguistics, we must not forget Luděk Hřebíček, Czech orientalist and quantitative linguist. He attempted to promote quantitative linguistics at time when it was not actually very popular (cf. e.g. Hřebíček, 2002, 2008). He then performed many experiments based on language samples of different languages, tested hypotheses walking the Altmanninian line (cf. Hřebíček, 1997, 2000, 2002, 2007). Nevertheless, he contributed with own well-processed ideas; to name at least some, we stress experimenting with suprasentence structures (he called them the aggregates), which drew him to other higher language structures above which there might remain only biological brain functions. He saw the language as a fractal entity so tended to go further than many of his contemporaries. He attempted, in his latest studies (cf. Hřebíček, 2007, 2007b), study semantics. This extraordinary attempt yet remained unfinished. He inspired many, we would at least like to mention works of the authors of this book, mentioned in the resource list.

The beginning of the 21st century brought a new wave into quantitative linguistics efforts. There appeared overlaps of quantitative linguistics and other, border or often non-linguistic sciences. Thus, we can even read works of fields outside linguistics utilizing quantitative linguistics methods. To name at least some, we mention the DNA analysis via quantitative linguistics methods, biosemiotics, psychology and psycholinguistics.

Manifestation of (not only) linguistic laws have been tested utilizing genetic texts (cf. Bolshoy, 2003). Some linguistic methods have traditionally been used in analyses of genetic texts e.g. on taxonomic purposes (Damerau–Levenshtein distance; cf. Damerau, 1964). In this scientific field, the research of linguistic laws is performed, an example can be employing the Zipf law for predicting the function of non-coding DNA (so called junk DNA; cf. Havlin et al., 1995). In this field of linguistic analyses of the genetic text, the researches of the Zipf law anticipate the research of the Menzerath–Altmann law and have a long-term tradition (cf. Niyogi, Berwick, 1995; Tsonis, Elsner, Tsonis, 1997). Universality of Zipf law manifestations in the text have been in the spotlight for long, see e.g. the question of law manifestations in a random text (cf. Ferrer-i-Cancho, Elvevag, 2010); an overview of the history of the current status of the Zipf law research and its prospects, including research overlaps outside the natural language are presented by Piantadosi (2014). Analyses of the Menzerath–Altmann and Zipf

laws still represent one of the most important linguistic approaches on the genetic text analysis, they have been used e.g. to describe the hierarchy of the genetic code and the protein structure (cf. Matlach, Faltýnek, 2016).

The Menzerath–Altmann law was examined in genetic texts with the following foci: Ferrer-i-Cancho et al. (2013) focus on the relation between the size and number of chromosomes in a genome (cf. Baixeries et al., 2013), also the exon sizes related to their number in a gene (cf. Nikolaou, 2014). Other researches monitor the relation of the size of protein domains and the protein sizes (cf. Shahzad, Mittenthal, Caetano-Anollés, 2015) or segments of coding and noncoding DNA (cf. Eroglu, 2014).

The Menzerath–Altmann law was, then, studied in the realm of animal communication. Manifestations of this law (and the Zipf law) are grabbed as an argument in a discussion on animal communication systems and their relation to the natural language (see so called language design features; Hockett, 1982). The law was studied on primates by Fedurek, Zuberbühler and Semple (2017), further by Gustison, Semple, Ferrer-i-Cancho and Bergman (2016).

The Menzerath–Altmann law was tested on texts by patients with language disorder. They were patients suffering from Broca's aphasia (cf. Jašíčková, Benešová, Faltýnek, 2013).

The above mentioned researches are often performed supposing manifestations of these laws are a natural language universal or a general structure. Related to animal language or speech disorders, they are understood as an instrument for language detection.

When we speak about quantitative linguistics, we must mention at least three journals issuing the outputs of the latest researches in this field. The leading journal is *Journal of Quantitative Linguistics*.
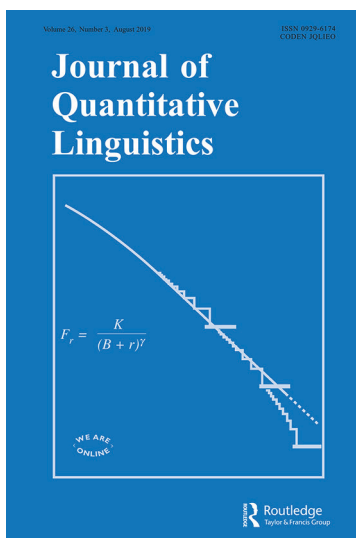

Figure 2.1: Front cover of *Journal of Quantitative Linguistics*

It is the official journal of the International Quantitative Linguistics Association and represents an international forum for the publication of work in the application of mathematics and statistics to linguistic research. It specifically publishes on modeling of all aspects of natural language in various relevant subdisciplines including theoretical linguistics, historical linguistics, sociolinguistics, dialectology, psycholinguistics and neurolinguistics; on practical application of mathematical and statistical methods in such areas as natural language processing, corpus linguistics, machine translation, information retrieval, and language teaching; on issues in the philosophy of language and linguistic theory of construction in the philosophy of science. For more information cf. www.iqla.org/iqla_journal.html. The journal blossomed for many years under the care of professor emeritus of Trier University Reinhard Köhler. It is worth mentioning that professor Köhler and professor Altmann were co-editors of The International Handbook of Quantitative Linguistics (cf. Köhler, Altmann, Piotrowski, 2005).

Another periodical of quantitative linguistics is *Glottometrics*. "The aim of Glottometrics is quantification, measurement and mathematical modeling of any kind of language phenomena. Any contributions on probabilistic or other mathematical models (e.g. graph theoretic or optimization approaches) which enable to establish language laws that can be validated by testing statistical hypotheses," for more information cf. www.ram-verlag.eu/journals-e-journals/glottometrics/aims-scope-editorial-board/. The editorial board was until recently lead also by professor Altmann.

Figure 2.2: Front cover of the journal *Glottometrics*

Following the use of quantitative methods in text analysis and the interdisciplinary cooperation of linguistics and natural sciences, the journal *Linguistic Frontiers* was created. The journal publishes articles on formal linguistics, quantitative linguistics, linguistic analysis of DNA, zoosemiotics, philosophy of linguistics etc.
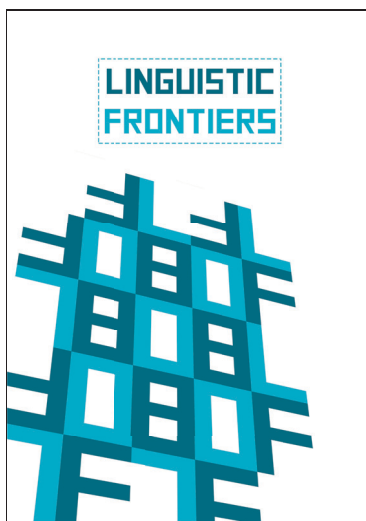
Figure 2.3: Front cover of the journal *Linguistic Frontiers*

Last but not least, *Glottotheory* should be mentioned. As the journal editors themselves proclaim, "the journal considers itself as platform for a dialogue between qualitative and quantitative linguistics", for more cf. www.degruyter.com/view/journals/glot/glot-overview.xml?lang=en.
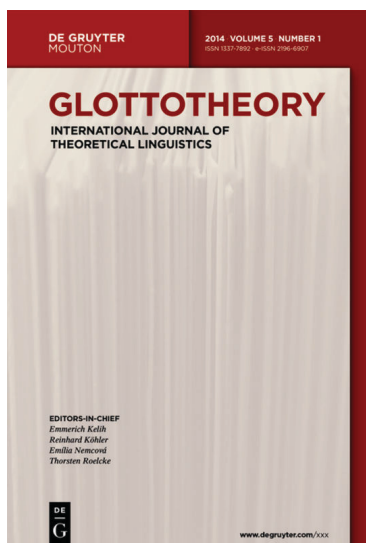
Figure 2.4: Front cover of the journal *Glottotheory*

The International Quantitative Linguistics association unites the researchers employing quantitative instruments or are interested in this field. It was established in 1994 "to promote the use of mathematical and statistical methods in

linguistic modeling, textual analysis and related fields. It is led by the IQLA Council, chaired by its President. IQLA issues the above mentioned journal and also organizes events and conferences of the community.

The already mentioned International Quantitative Linguistics Association organizes also every other year regular conferences under the name of "QUALICO". The QUALICO conference in the recent decade have been held in Belgrade, Serbia (2012), Olomouc, Czech Republic (2014), Trier, Germany (2016), Wroclaw, Poland (2018). For more information on QUALICO conferences, cf. www.iqla.org/iqla_conferences.html.

2.2 Menzerath–Altmann law (MAL)

MAL is one ot the key laws of *quantitative linguistics* which describes an inverse proportion between the integer length x of *constructs* (i.e. the language units of higher levels) and the averaged length y of *constituents* (i.e. the language units of lower levels), composing the constructs. Its verbal version reads "the longer construct is, the shorter constituents are." Let us note that the lengths are usually measured in the number of units of the closest lower language unit. An alternative possibility of measuring the lengths is in seconds. The most sophisticated way is related to measuring the lengths of parameters at simultaneities.

The *complete formula of MAL* takes the form

$$y = A\,x^{-b}\mathrm{e}^{cx},$$

where A, b, c are real constants and $\mathrm{e} = 2.718\dots$ is the Euler number.

For $c = 0$, we obtain the *truncated formula of MAL*, namely

$$y = Ax^{-b}.$$

In particular, for $A = y(1) = y_1$, the (reduced) truncated formula contains only one free constant b, which is sometimes called a *shape parameter*.

By means of *MAL*, we can formulate a *conjecture about the fractal structure of spoken languages* and, on its basis, to define *language fractals*. It is a question whether or not one can extend this conjecture to a *sign language fractals*.

2.3 Zipf–Mandelbrot law

The *Zipf–Mandelbrot law (ZML)* describes the distribution of word (sign) frequencies for regular lexicographic trees. It takes the mathematical form

$$U = P(\rho + V)^{-\frac{1}{D}}$$

or, equivalently,

$$D = \frac{\log(\rho + V)}{\log \frac{P}{U}},$$

where ρ denotes the order of a word (sign), ordered in a decreasing way according to the related frequencies, with probability U, and P, V, D are real constants. The

regularity of trees means that each branching is related to a single word (sign) and the probability weight on the k-th level takes the form $U = U_0 r^k$, where $0 < r < 1$, and U_0 is such that the sum of all probability weights is equal to 1.

The restriction can be relaxed for cyclically (i.e. block-like) self-similar trees, like for the Menzerath–Altmann law, where this time

$$D := \frac{m \log(\rho + V)}{\sum_{j=1}^{m} \log \frac{P_j}{U_j}},$$

resp.

$$\rho = -V + \exp\left(\frac{D}{m} \sum_{j=1}^{m} \ln \frac{P_j}{U_j}\right).$$

There exist also further various versions of the Zipf law, like the *Zipf–Alekseev–Dolinskii law*, etc.

2.4   Conjecture about the fractal structure of language

By a *language fractal*, one can sometimes understand the *semantic recursions*, where several meta-levels can be distinguished. Many such expressions can be found in poetry, for instance: "the abysm of an abysm" (V. Holan), "I know … what I know" (W. Stevens), etc.

Mandelbrot was probably the first who studied systematically self-similar language structures and their fractality since the late 60's. For this aim, he investigated the *regular lexicographical trees* by means of the Zipf–Mandelbrot law, where "each branch represents itself the scaled reduced version of the whole tree". Although the lexicographical trees are in general neither regular, nor self-similar, Mandelbrot's ideas in this field can be regarded as an initial stimulation for *fractal analysis of natural spoken languages*.

Altmann's concept of quantitative linguistics (see e.g. Altmann, 2002, 2012), lead Hřebíček to formulate a *conjecture about the fractal structure of spoken languages*. Its *weak version* says that *the Menzerath–Altmann law (MAL) holds for known language levels*, including supra-sentence levels like *semantic constructs*. In other words that *(spoken) language structures exhibit self-similarity with respect to the validity of MAL on all language levels*, but up to constant parameters A, b, c by which a given language structure is characterized.

The *strong version* is based on an isomorphism between logarithmized form of MAL and the *Moran–Hutchinson formula* for the calculation of a *similarity (fractal) dimension* of a mathematical fractal. According to this version, we can associate to a given language structure, satisfying MAL with positive shape parameters b, on all levels under consideration, a cyclically (i.e. block-like) self-similar mathematical fractal whose dimension equals the reciprocal value of an arithmetic mean of respective shape parameters $b > 0$.

2.5   Language fractals

In view of the correspondence

$$D_s \sim \frac{1}{b}, \quad m \sim x, \quad r \sim \frac{y}{A\exp(cx)} = x^{-b},$$

between (for $b \neq 0$) an equivalent form of the *Menzerath–Altmann law (MAL)*, i.e.

$$\frac{1}{b} = \frac{\log x}{\log\left(\frac{A}{y}\exp(cx)\right)} = \frac{\ln x}{\ln\left(\frac{A}{y}\exp(cx)\right)},$$

and the *Moran–Hutchinson formula* for the calculation of the similarity (fractal) dimension of a mathematical fractal, i.e.

$$D_s = \frac{\log m}{\log\frac{1}{r}},$$

where m denotes the number of its components on each scale and $r \in [0,1)$ is a related contraction factor, we can give the following definition of a language fractal, provided the model fractal set is either totally disconnected or its components are just touching. Then by a *language fractal* we understand a language structure satisfying MAL with positive shape parameters b on all language levels under consideration. In this case, the mathematical model approximates a certain cyclically (i.e. block-like) self-similar fractal, whose dimension D equals the reciprocal value of an arithmetic mean of respective shape parameters $b > 0$.

Thus, if we consider n language levels distinguished by the indices $i = 1, 2, \ldots, n$, with the respective shape parameters $b_i > 0$, $i = 1, 2, \ldots, n$, then the value $D := n/(b_1 + \cdots + b_n)$ will be called (in view of the above correspondence) the *measure of semanticity* of a given language fractal, resp. the dimension of an associated cyclically self-similar modeling fractal.

The *fractal analysis of texts* therefore allows us to define their measures of semanticity and to visualize the models of given language structures.

2.6   Mathematical fractals

Mathematical fractals are, roughly speaking, infinitely broken geometrical objects. Despite the non-existence of a unique mathematical definition, by a mathematical fractal, we usually mean at least one of the following possibilities:

  (i) a geometrical object, whose *(noninteger) fractal dimension* is strictly higher than its (integer) topological dimension (definition of Mandelbrot);

 (ii) a geometrical object with an axiomatically postulated property of *self-similarity*, where the whole object is similar to its parts on given scales (see e.g. Feder, 1988);

(iii) an *attractor of the iterated function system (IFS)* in the frame of discrete dynamical systems (definition of Hutchinson, see e.g. Barnsley, 2013).

These definitions do not coincide each to other in general which could bring confusion, when we do not respect the specific differences. On the other hand, the majority of standard (classical) fractals satisfy all the properties (i)–(iii), i.e. that they belong to their intersection.

Although mathematical fractals do not exist in nature, they are suitable for modeling natural shapes and forms, like for the morphological description of plants (e.g. Barnsley's fern), but also for the characterization of a fractal structure of languages. In the latter sense, we speak about *language fractals*.

Figure 2.5: Barnsley's fern

Unlike to the classical fractals (e.g. Cantor's dust, von Koch's flake, Sierpiń-ski's triangle), which satisfy (i)–(iii), for language fractals we usually require only some of the properties, like (iii), resp. (ii), but not necessarily (i).
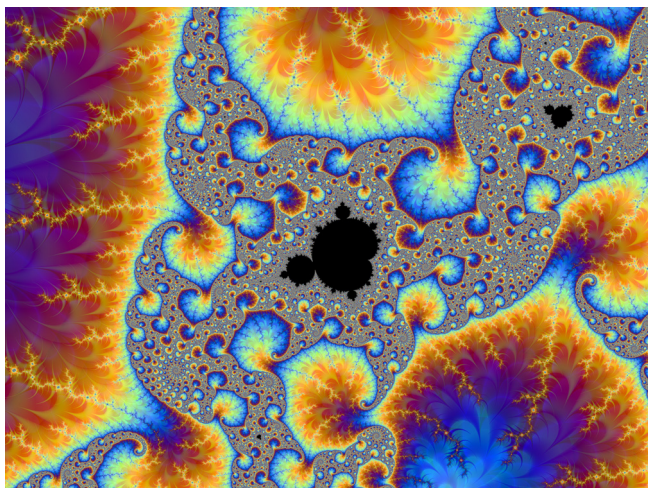
Figure 2.6: Some mathematical fractals are beautiful (adapted from Wikipedia Commons, free license)

2.7 Hurst exponent vs. fractal dimension

The (asymptotic) Hurst formula takes the form

$$E\left[\frac{R(\tau)}{S(\tau)}\right] = C\tau^{H}, \quad \text{as } \tau \to \infty, \ (C \geq 0 \text{ is a constant}),$$

where the constant $H \geq 0$ is called the *Hurst* (sometimes also *Hurst–Hölder*) *exponent*, the letter "E" denotes the expected value of the ratio $R(\tau)/S(\tau)$ depending on the increasing number τ of data points in a given time series $x_1, x_2, \ldots, x_\tau,$ $\ldots, x_N$.

- $R(\tau)$ denotes the range of the first τ cumulative deviations of the mean

$$\bar{x}(\tau) = \frac{1}{\tau}\sum_{i=1}^{\tau} x_i,$$

  i.e.

$$R(\tau) := \max_{1 \leq t \leq \tau} x(t, \tau) - \min_{1 \leq t \leq \tau} x(t, \tau), \quad x(t, \tau) := \sum_{i=1}^{t}(x_i - \bar{x}(\tau)), \quad t = 1, 2, \ldots, \tau,$$

- $S(\tau) := \sqrt{\dfrac{1}{\tau}\sum_{i=1}^{t}(x_i - \bar{x}(\tau))^2}$ denotes the standard deviation.

  For the *"finite time"* *Hurst formula*

$$\frac{R(\tau)}{S(\tau)} = C_\tau \tau^{H_\tau}, \quad \text{where } S(\tau) > 0,$$

the related Hurst exponent $H_\tau = H(\tau)$ can be easily calculated as $(\tau > 1)$

$$H_\tau = \frac{\log \frac{1}{C_\tau}\frac{R(\tau)}{S(\tau)}}{\log \tau}.$$

- For self–similar time series, describing a one–dimensional process, we can define the *box–counting dimension D* as

$$D_{bc} = 2 - H, \quad \text{where } D_{bc} \in [1, 2].$$

- For (not necessarily one–dimensional) self-affine structures, we can also define the *divider dimension D* as

$$D_d = \frac{1}{H}.$$

- For cyclically (i.e. block-like) self-similar m families of time series, we can define the *Hausdorff dimension D_H* as the reciprocal mean value of the respective Hurst exponents $H_1, H_2, \ldots, H_m$ (on m scaling levels), i.e.

$$D_H = \frac{m}{\sum_{j=1}^{m} H_j},$$

  provided the "open set condition" (i.e., in particular, with totally disconnected or at most just bounding components) is satisfied.

The critical value is $H = 0.5$, indicating no correlation in a given time series. For $H > 0.5$, a time series exhibits a long-term positive autocorrelation (persistence). Otherwise, i.e. for $H < 0.5$, an erratical oscillation occurs in a time series.

2.8   Lyapunov exponents

The standard Lyapunov exponent is an important indicator for measuring a sensitive dependence of dynamics on initial values. Roughly speaking, the Lyapunov exponent measures at each point the propagation of errors after each iteration of a given function or sequence. In other words, it measures the mean loss of information during the iteration process at the points which are close to the initial values.

Let $(y_0 = 0), y_1, y_2, \ldots, y_n$ be the sequence of lengths of (sign) language units. Then the local Lyapunov-like exponents λ_j, related to this sequence, can be simply defined (in a somewhat different sense as above) as follows:

$$\lambda_j := \frac{1}{j} \ln \left| \frac{y_j - y_{j+1}}{y_0 - y_1} \right| = \frac{1}{j} \sum_{i=1}^{j} \ln \left| \frac{y_i - y_{i+1}}{y_{i-1} - y_i} \right|, \quad j = 1, \ldots, n-1.$$

In order to have a global formula in the above sense, we can apply splines of degree k. Thus, the sequence $\{y_i\}_{i=0}^{n}$ can be smoothly approximated by the polynomial function P_k. In this way the "discrete" local formula can be replaced by the *"continuous" formula* for P_k, namely

$$\lambda(x_0) \approx \frac{1}{m} \sum_{j=l+1}^{l+m} \ln \left| P_k'(x_j) \right|,$$

where $x_0 \in (0, n)$, $x_j := P_k^j(x_0) = \underbrace{P_k \circ \ldots \circ P_k}_{j\text{-times}}(x_0)$ is the j-th iteration of P_k at x_0, and l, m are sufficiently big numbers (e.g. $l = 400$, $m = 500$).

The *exact continuous formula* for the Lyapunov exponent $\lambda(x_0)$ to P_k takes the form

$$\lambda(x_0) := \lim_{l \to \infty} \frac{1}{l} \sum_{j=0}^{l-1} \ln \left| P_k'(x_j) \right|.$$

Since the second formula is a "numerical" approximation of the last one, it is preferable for the practical usage. The critical value for the Lyapunov exponent is zero; negative values indicate order, while positive values indicate instability which can lead to chaos.

Peter Angermann, *Noch ein Homunculus*, 1989, crayon

3 MENZERATH–ALTMANN LAW (MAL)

This chapter can be regarded as a more advanced approach to the previous entry.

3.1 Derivation of formulas for MAL

The *Menzerath–Altmann law* (shortly, MAL) is with no doubts one of the milestones of quantitative linguistics. Its verbal (heuristic) form says that "the longer a language construct is, the shorter its components (constituents) are." By a construct we mean a unit on a higher language level (e.g. a *clause*, i.e. an autonomous unit in terms of its pragmatics, semantic construction and grammar which is based on a finite verb form) and by a constituent we mean a unit on a lower language level (e.g. a word, resp. a sign). The length x of a construct is measured in its constituents, while y is the average length of its constituents measured in units on the nearest lower language level. For instance, x can be a length of a clause measured in the number of words (resp. in a sign language, in the number of signs) and y then denotes the average number of composing words measured in the number of syllables.

Although the length x of a construct is (unlike $y \in (0, \infty)$) usually a positive integer, i.e. $x \in \mathbb{N}$, for the sake of derivation of the mathematical "continuous" form of MAL, we will nevertheless consider x to be also positively real-valued, i.e. $x \in (0, \infty)$.

Hence, recalling the mathematical derivation of MAL due to Gabriel Altmann (cf. Altmann, 1980; Altmann, Schwibbe, Kaumanns, 1989; Wimmer, Altmann, 2005), the relative rate of the change of y can be expressed as $\dot{y}/y$, where $\dot{y} = \mathrm{d}y/\mathrm{d}t$ is obviously the derivative of y w.r.t. x. According to the verbal formulation above, this rate is inversely proportional to x up to an additive constant $c \in \mathbb{R}$, i.e.

$$\frac{\dot{y}}{y} = \frac{-b}{x} + c, \tag{3.1}$$

where $-b$ is the proportionality coefficient. Let us note that it is convenient to take it with the minus sign, as we shall see later.

Integrating the ordinary differential equation (3.1) w.r.t. x, we obtain the equation $\ln y = -b \ln x + cx + C$, where $C \in \mathbb{R}$ is another additive constant, i.e. after delogarithmization, we get

$$y = \mathrm{e}^{C} x^{-b} \mathrm{e}^{cx}.$$

Thus, the general solution of (3.1) takes the form

$$y(x) = Ax^{-b}e^{cx}, \quad \text{where } A = e^C > 0. \tag{3.2}$$

This is so called a *complete formula* of MAL.

Let us note that in many empirical studies concerning sentence or clause structures, and all the better in supra-sentence structures like semantic constructs, only its hyperbolic part is used, namely ($c = 0$)

$$y = Ax^{-b}, \tag{3.3}$$

(see e.g. Hřebíček, 1995, 2007). Here, we therefore speak about the *truncated formula* of MAL.

On the other hand, it was demonstrated that the role of the exponential parts, which can be omitted in the case of semiotically higher levels, increases with a decreasing linguistic level, and so it is usually not omitted (for instance, in the case of words and syllables).

In order to simplify the computation of parameters b, c in (3.2) or b in (3.3), we can still put $A := y(1)/e^c = y_1/e^c$ in (3.2), i.e.

$$y = y_1 x^{-b} e^{c(x-1)}, \tag{3.4}$$

or $A := y(1) = y_1$ in (3.3), i.e.

$$y = y_1 x^{-b}. \tag{3.5}$$

How much the formulas (3.2)–(3.5) differ each from other? For the first glance, it might be expected that the optimal results must be connected with the complete formula (3.2), while the roughest are those related to the truncated formula (3.5). However, the situation is not so simple but rather delicate for such a statement.

For more details concerning various aspects of MAL, we recommend e.g. the papers Altmann (1980), Altmann, Schwibbe, Kaumanns (1989), Andres (2009, 2010), Cramer (2005), Hřebíček (1990, 1992, 1995, 2007), Köhler (1982, 1984, 1989), Kulacka (2010), Kulacka, Mačutek (2007), Prün (1994), Wimmer, Altmann (2005) and the book Wimmer, Altmann, Hřebíček, Ondrejovič, Wimmerová (2003). In Wimmer, Altmann (2005), the whole panorama of further variants of linguistic formulas of this sort is presented.

3.2   Accuracy of approximations of measured data (least square method)

Let us consider separately the following cases:

   I) $y = y_1 x^{-b}$, i.e. $A = y(1) = y_1$ and $c = 0$, (cf. (3.5)),

  II) $y = Ax^{-b}$, i.e. $c = 0$, (cf. (3.3)),

 III) $y = y_1 x^{-b} e^{c(x-1)}$, i.e. $Ae^c = y(1) = y_1$, (cf. (3.4)),

 IV) $y = Ax^{-b} e^{cx}$, (cf. (3.2)).

ad I) (one free parameter b)

Logarithmizing the equation (3.5), we obtain

$$\ln y = \ln y_1 - b \ln x,$$

i.e.

$$\ln y_i = \ln y_1 - b \ln x_i, \quad i = 1, 2, \dots,$$

where $x_i = i$ denotes the length of the i-th construct and y_i denotes the length of the i-th constituent of x_i.

Hence, denoting still $u_i := \ln x_i$ and $v_i := \ln y_i$, we can minimize the function

$$\psi_I(b) := \sum_i w_i [v_1 - b u_i - v_i]^2$$

w.r.t. b, where $w_i = z_i / \sum_i z_i$ are the weights corresponding to the i-th relative frequency $x_i / \sum_i z_i$.

Putting

$$\psi_I(b) = 2 \sum_i w_i [v_1 - b u_i - v_i](-u_i) = 0,$$

we arrive at

$$b_I = \frac{v_1 \sum_i w_i u_i - \sum_i w_i u_i v_i}{\sum_i w_i u_i^2} = \frac{\ln y_1 \sum_i w_i \ln x_i - \sum_i w_i \ln x_i \ln y_i}{\sum_i w_i (\ln x_i)^2}. \tag{3.6}$$

Since

$$\ddot{\psi}_I(b) := 2 \sum_i w_i u_i^2 > 0,$$

we really have that

$$\min_b \psi_I(b) = \psi_I(b_I).$$

Moreover, the related least square value Δ_I takes the form

$$\Delta_I := \sum_i w_i [y(x_i) - y_i]^2 = \sum_i w_i \left[y_1 x_i^{-b_I} - y_i \right]^2. \tag{3.7}$$

Remark 3.1. Let us note that b_I can slightly differ from $\hat{b}_I$ satisfying

$$\min_b \sum_i w_i \left[y_1 x_i^{-b_I} - y_i \right]^2 = \sum_i w_i \left[y_1 x_i^{-\hat{b}_I} - y_i \right]^2,$$

but statistically b_I, and subsequently also Δ_I, can be acceptable (see e.g. Kubáček, Kubáčková, 2000, pp. 219–225).

ad II) (two free parameters A, b)

Logarithmizing the equation (3.3), we obtain

$$\ln y = \ln A - b \ln x,$$

i.e.

$$\ln y_i = \ln A - b \ln x_i, \quad i = 1, 2, \ldots,$$

where the symbols x_i and y_i have the same meaning as above.

Hence, denoting again $u_i := \ln x_i$, $v_i := \ln y_i$ and newly $a := \ln A$, we can minimize the function

$$\psi_{II}(a,b) := \sum_i w_i[a - bu_i - v_i]^2$$

w.r.t. $a = \ln A$, b, where the weights $w_i = z_i / \sum_i z_i$ are the same as in the case I).

Putting

$$\frac{\partial \psi_{II}}{\partial a} = 2\sum_i w_i[a - bu_i - v_i] = 0, \qquad \frac{\partial \psi_{II}}{\partial b} = 2\sum_i w_i[a - bu_i - v_i](-u_i) = 0,$$

we arrive at

$$A_{II} = \exp\left(\frac{\sum_i w_i \ln x_i \sum_i w_i (\ln x_i)(\ln y_i) - \sum_i w_i (\ln x_i)^2 \sum_i w_i \ln y_i}{[\sum_i w_i \ln x_i]^2 - \sum_i w_i (\ln x_i)^2}\right), \qquad (3.8)$$

$$b_{II} = \frac{\sum_i w_i (\ln x_i)(\ln y_i) - \sum_i w_i \ln x_i \sum_i w_i \ln y_i}{[\sum_i w_i \ln x_i]^2 - \sum_i w_i (\ln x_i)^2}. \qquad (3.9)$$

Since

$$\frac{\partial^2 \psi_{II}}{\partial a^2} = 2\sum_i w_i = 2 > 0$$

and (in view of the Schwarz inequality)

$$\frac{\partial^2 \psi_{II}}{\partial a^2}\frac{\partial^2 \psi_{II}}{\partial b^2} = \left[\frac{\partial^2 \psi_{II}}{\partial a \partial b}\right]^2 = 4\sum_i w_i u_i^2 - 4\left[\sum_i w_i u_i\right]^2 > 0,$$

we really have that

$$\min_{a,b} \psi_{II}(a,b) = \psi_{II}(\ln A_{II}, b_{II}).$$

Moreover, the related least square value Δ_I takes the form

$$\Delta_{II} := \sum_i w_i \left[A_{II}x_i^{-b_{II}} - y_i\right]^2. \qquad (3.10)$$

ad III) (two free parameters b, c)

Logarithmizing the equation (3.4), we obtain

$$\ln y = \ln y_1 - b \ln x + c(x - 1),$$

i.e.

$$\ln y_i = \ln y_1 - b \ln x_i + c(x_i - 1), \quad i = 1, 2, \ldots,$$

where the symbols x_i and y_i have the same meaning as above.

Hence, denoting again $u_i := \ln x_i (\Rightarrow x_i = \exp(u_i))$ and $v_i = \ln y_i$, we can minimize the function

$$\psi_{III}(a, b) := \sum_i w_i [v_1 - b u_i + c(x_i - 1) - v_i]^2$$

w.r.t. b, c, where the weights $w_i = z_i / \sum_i z_i$ are the same as above.

Putting

$$\frac{\partial \psi_{III}}{\partial b} = 2 \sum_i w_i [v_1 - b u_i + c(x_i - 1) - v_i](-u_i) = 0,$$

$$\frac{\partial \psi_{III}}{\partial c} = 2 \sum_i w_i [v_1 - b u_i + c(x_i - 1) - v_i](x_i - 1) = 0,$$

we arrive at $(A_{III} = y_1 / e_{III}^c)$

$$b_{III} = \frac{\sum_i w_i (x_i - 1)^2 \sum_i w_i \ln x_i (\ln y_1 - \ln y_i)}{\sum_i w_i (\ln x_i)^2 \sum_i w_i (x_i - 1)^2 - [\sum_i w_i (\ln x_i) \ln x_i]^2}$$
$$+ \frac{\sum_i w_i (x_i - 1) \ln x_i \sum_i w_i (x_i - 1)(\ln y_1 - \ln y_i)}{\sum_i w_i (\ln x_i)^2 \sum_i w_i (x_i - 1)^2 - [\sum_i w_i (\ln x_i) \ln x_i]^2}, \quad (3.11)$$

$$c_{III} = \frac{\sum_i w_i (x_i - 1) \ln x_i \sum_i w_i \ln x_i (\ln y_1 - \ln y_i)}{\sum_i w_i (\ln x_i)^2 \sum_i w_i (x_i - 1)^2 - [\sum_i w_i (\ln x_i) \ln x_i]^2}$$
$$+ \frac{\sum_i w_i (\ln x_i)^2 \sum_i w_i (x_i - 1)(\ln y_1 - \ln y_i)}{\sum_i w_i (\ln x_i)^2 \sum_i w_i (x_i - 1)^2 - [\sum_i w_i (x_i - 1) \ln x_i]^2}. \quad (3.12)$$

Since

$$\frac{\partial^2 \psi_{III}}{\partial b^2} = 2 \sum_i w_i u_i^2 = 2 \sum_i w_i (\ln x_i)^2 > 0$$

and (in view of the Schwarz inequality)

$$\frac{\partial^2 \psi_{III}}{\partial b^2} \frac{\partial^2 \psi_{III}}{\partial c^2} - \left[\frac{\partial^2 \psi_{III}}{\partial b \partial c} \right]^2$$
$$= 4 \sum_i w_i (\ln x_i)^2 \sum_i w_i (x_i - 1)^2 - 4 \left[\sum_i w_i (x_i - 1) \ln x_i \right]^2 > 0,$$

we really have that

$$\min_{b,c} \psi_{III}(b,c) = \psi_{III}(b_{III},c_{III}).$$

Moreover, the related least square value Δ_{III} takes the form

$$\Delta_{III} := \sum_i w_i \left[y_1 x_i^{-b_{III}} e^{c_{III}(x_i-1)} - y_i \right]^2. \tag{3.13}$$

ad IV) (three free parameters A, b, c)

Logarithmizing the equation (3.7), we obtain

$$\ln y = \ln A - b \ln x + cx,$$

i.e.

$$\ln y_i = \ln A - b \ln x_i + c x_i, \quad i = 1, 2, \ldots,$$

where the symbols x_i and y_i have the same meaning as above.

Hence, denoting again $u_i := \ln x_i (\Rightarrow x_i = \exp(u_i))$ and $v_i = \ln y_i$ and $a := \ln A$, we can minimize the function

$$\psi_{IV}(a,b,c) := \sum_i w_i [a - bu_i + cx_i - v_i]^2$$

w.r.t. a, b, c, where the weights $w_i = z_i / \sum_i z_i$ are the same as above.

Putting

$$\frac{\partial \psi_{IV}}{\partial a} = 2 \sum_i w_i [a - bu_i + cx_i - v_i] = 0,$$

$$\frac{\partial \psi_{IV}}{\partial b} = 2 \sum_i w_i [v_1 - bu_i + cx_i - v_i](-u_i) = 0,$$

$$\frac{\partial \psi_{IV}}{\partial c} = 2 \sum_i w_i [v_1 - bu_i + cx_i - v_i] x_i = 0,$$

we arrive at $(A_{IV} = \exp(a_{IV}))$

$$a_{IV} = \frac{1}{\tau} \left\{ \sum_i w_i \ln y_i \left[\left(\sum_i w_i x_i \ln x_i \right)^2 - \sum_i w_i x_i^2 \sum_i w_i (\ln x_i)^2 \right] \right.$$

$$+ \sum_i w_i \ln x_i \ln y_i \left[\sum_i w_i x_i^2 \sum_i w_i \ln x_i - \sum_i w_i \ln x_i \sum_i w_i x_i \ln x_i \right]$$

$$+ \left. \sum_i w_i x_i \ln y_i \left[\sum_i w_i x_i \sum_i w_i (\ln x_i)^2 - \sum_i w_i \ln x_i \sum_i w_i x_i \ln x_i \right] \right\}, \tag{3.14}$$

$$b_{IV} = \frac{1}{\tau} \left\{ \sum_i w_i \ln y_i \left[\sum_i w_i x_i \sum_i w_i x_i \ln x_i - \sum_i w_i x_i^2 \sum_i w_i \ln x_i \right] \right.$$

$$+ \sum_i w_i \ln x_i \ln y_i \left[\sum_i w_i x_i^2 - \left(\sum_i w_i x_i \right)^2 \right]$$

$$\left. + \sum_i w_i x_i \ln y_i \left[\sum_i w_i x_i \sum_i w_i \ln x_i - \sum_i w_i x_i \ln x_i \right] \right\}, \quad (3.15)$$

$$c_{IV} = \frac{1}{\tau} \left\{ \sum_i w_i \ln y_i \left[\sum_i w_i x_i \sum_i w_i (\ln x_i)^2 - \sum_i w_i \ln x_i \sum_i w_i x_i \ln x_i \right] \right.$$

$$+ \sum_i w_i \ln x_i \ln y_i \left[\sum_i w_i x_i \ln x_i - \sum_i w_i x_i \sum_i w_i \ln x_i \right]$$

$$\left. + \sum_i w_i x_i \ln y_i \left[\left(\sum_i w_i \ln x_i \right)^2 - \sum_i w_i (\ln x_i)^2 \right] \right\}, \quad (3.16)$$

where

$$\tau = \left(\sum_i w_i x_i \ln x_i \right)^2 + \left(\sum_i w_i \ln x_i \right)^2 \sum_i w_i x_i^2 + \left(\sum_i w_i x_i \right)^2 \sum_i w_i (\ln x_i)^2$$

$$- \sum_i w_i (\ln x_i)^2 \sum_i w_i^2 - 2 \sum_i w_i \ln x_i \sum_i w_i x_i \sum_i w_i x_i \ln x_i.$$

We can namely prove again (see e.g. Kubáček, Kubáčková, 2000, p. 83) that

$$\min_{a,b,c} \psi_{IV}(a,b,c) = \psi_{IV}(a_{IV}, b_{IV}, c_{IV}).$$

Moreover, the related least square value Δ_{IV} takes the form

$$\Delta_{IV} := \sum_i w_i \left[A_{IV} x_i^{-b_{IV}} e^{c_{IV} x_i} - y_i \right]^2. \quad (3.17)$$

Summing up the above calculations, we can give the following conclusion.

Conclusion 3.1. *Approximative values of parameters A, b, c in formulas (3.2)–(3.5) of MAL can be obtained from the lengths of associated constructs $x_1, x_2, \ldots$, their frequencies $z_1, z_2, \ldots$, and constituents $y_1, y_2, \ldots$, by means of a regression method in the forms (3.6) and (3.8), (3.9) and (3.1), (3.12) and (3.14)–(3.9), respectively. The appropriate least square values take the forms (3.7) and (3.10) and (3.13) and (3.17), respectively.*

3.3   Comparison of accuracy of parameter estimations

In order to compare formulas (3.2)–(3.5), the experiment design theory should be employed (see e.g. Pázman, 1988). This technique is closely related to the previous one but the output is different. It does not concern the information about conformity of data with the applied approximation formula, but the value of variance of the free parameters estimates (in statistical terms, we speak about unknown parameters) $\hat{\Theta}_I := b$ or $\hat{\Theta}_{II} := (A,b)'$ or $\hat{\Theta}_{III} := (b,c)'$ or $\hat{\Theta}_{IV} := (A,b,c)'$ which appear in the respective formulas (3.2)–(3.5). The prime symbol denotes as usually the transposition. Although this analysis is again based on the least square method, it is necessary to introduce adequate terms because of input variables constructed in a different way. After logarithmizing the equations (3.2)–(3.5), we obtain four linear regression models. Their general form is

$$y \sim_n (\boldsymbol{F}\boldsymbol{\Theta}, \sigma^2, \boldsymbol{I}), \tag{3.18}$$

where y has the same meaning as above, n is the number of measurement points, $\boldsymbol{F}$ is a *design matrix* given by the concrete formula (3.2) or (3.3) or (3.4) or (3.5), $\boldsymbol{\Theta}$ is an appropriate *vector of unknown parameters*, σ^2 is a *measurement error* and $\boldsymbol{I}$ denotes the *unit matrix*. We assume the independence of random variables y_i, $i = 1, \ldots, n$.

Without any loss of generality, we can now restrict ourselves to a specific area of the experiment realization which is known as the *set of experimental points* $x = x_1, \ldots, x_n$. The symbol $x_i = i$ has the same meaning as above and in order to guarantee the sufficient richness of the experimental set, we consider the total number of measurements n to be equal to $10\,000$.

The following analysis is based on the validity of the formula (see Kubáček, Kubáčková, 2000, p. 50)

$$\mathrm{Var}(\hat{\boldsymbol{\Theta}}(\boldsymbol{Y}_\delta), N) := \frac{\sigma^2}{N} \boldsymbol{M}^{-1}(\delta), \tag{3.19}$$

whose proof can be found in Tučková (2010). The symbol N denotes the total number of measurements at the points x_i, i.e. $N = \sum_i z_i$. It says that the covariance matrix estimate of unknown parameters (contained in a regression model) is equal to the product of the expression $\frac{\sigma^2}{N}$ and the inverted matrix $\boldsymbol{M}(\delta)$, called *information matrix*, namely

$$\boldsymbol{M}(\delta) := \sum_{i=1}^{n} \delta_i w_i \boldsymbol{f}_i \boldsymbol{f}_i'. \tag{3.20}$$

The form (3.20) of an information matrix is given by several variables. In particular, the *design of experiment* is a function

$$\delta_i \colon x_i \to [0,1], \quad \text{where } \sum_{i=1}^{n} \delta_i = 1, \; i = 1, \ldots, n,$$

which determines a significance of the experiment at various points of the experimental set. In our case, the significance is the same for all experimental points.

Therefore, the value of design of experiment is a constant, i.e. $\delta_i = \frac{1}{n}$. Variable $w_i = z_i / \sum_i z_i$ represents the above mentioned weight. Since this value is constant in formulas (3.2)–(3.5), it can be neglected, without any loss of generality, in subsequent calculations. Last, but not least, vector $\boldsymbol{f}_i$ is the variable which represents i-th line of design matrix $\boldsymbol{F}$.

Expression $\frac{\sigma^2}{N}$ in ((3.19) depends on the particular experiment, but it is the same for all models in (3.18). Thus, this expression can be put equal to 1, for the sake of comparison of formulas (3.2)–(3.5).

The comparison of regression models will be made by means of a *standard deviation* of individual unknown parameter estimates, i.e.

$$\mu(\hat{\boldsymbol{\Theta}}) := \sqrt{\mathrm{diag}\left[\mathrm{Var}(\hat{\boldsymbol{\Theta}})\right]}. \tag{3.21}$$

ad I) (one free parameter b)

Logarithmizing the equation (3.5), we arrive in view of (3.18) to the equality:

$$\begin{pmatrix} 0 \\ v_2 - \ln y_1 \\ \vdots \\ v_n - \ln y_1 \end{pmatrix} = \underbrace{\begin{pmatrix} -u_1 \\ \vdots \\ -u_n \end{pmatrix}}_{\boldsymbol{F}} (b) + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}$ is an error vector with the covariance matrix equal to $\sigma^2 \boldsymbol{I}$. The variables y_i, u_i and v_i have the same meaning as above.

The information matrix takes, according to (3.20) (when taking the weights $w_i = 1$), the form

$$\boldsymbol{M}(\delta) = \sum_{i=1}^{n} \frac{1}{n}(-u_i)(-u_i) = \frac{1}{n} \sum_{i=1}^{n} u_i^2,$$

from which (cf. (3.19), when taking $\frac{\sigma^2}{N} = 1$)

$$\mathrm{Var}_I(\hat{\boldsymbol{\Theta}}(\boldsymbol{Y}_\delta)) \doteq 1.461694 \cdot 10^{-6},$$

and subsequently (see (3.6))

$$\mu(b_I) \doteq \sqrt{1.461694 \cdot 10^{-6}} \doteq 0.001209. \tag{3.22}$$

ad II) (two free parameters A, b)

Logarithmizing the equation (3.3), we arrive in view of (3.18) to the equality:

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \underbrace{\begin{pmatrix} 1, & -u_1 \\ & \vdots \\ 1, & -u_n \end{pmatrix}}_{\boldsymbol{F}} \begin{pmatrix} a \\ b \end{pmatrix} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}$ has the same meaning as in the foregoing case. The variables y_i, u_i and v_i have the same meaning as above and $a = \ln A$.

The information matrix takes, according to (3.20) (when taking the weights $w_i = 1$), the form

$$\boldsymbol{M}(\delta) = \sum_{i=1}^{n} \frac{1}{n} \begin{pmatrix} 1 \\ -u_i \end{pmatrix} (1, -u_i) = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1, & -u_i \\ -u_i, & u_i^2 \end{pmatrix} = \frac{1}{n} \begin{pmatrix} n, & -\sum_{i=1}^{n} u_i \\ -\sum_{i=1}^{n} u_i, & \sum_{i=1}^{n} u_i^2 \end{pmatrix},$$

from which (cf. (3.19), when taking $\frac{\sigma^2}{N} = 1$)

$$\mathrm{Var}_{II}(\hat{\boldsymbol{\Theta}}(\boldsymbol{Y}_\delta)) \doteq \begin{pmatrix} 0.006875, & 0.000825 \\ 0.000825, & 0.000100 \end{pmatrix},$$

and subsequently (see (3.8), (3.9))

$$\mu(\ln A_{II}) \doteq \sqrt{0.006875} \doteq 0.082921, \tag{3.23}$$

i.e. (cf. Kubáček, Kubáčková, 2000, p. 215)

$$\mu(A_{II}) = \frac{\partial \mathrm{e}^a}{\partial a} \Big|_{a = \ln A_{II}} \mu(\ln A_{II}) = A_{II} \mu(\ln A_{II})$$
$$\doteq 2.0815 \cdot 0.08291 \doteq 0.170919, \tag{3.24}$$

$$\mu(b_{II}) \doteq \sqrt{0.000100} \doteq 0.010025. \tag{3.25}$$

ad III) (two free parameters b, c)

Logarithmizing the equation (3.4), we arrive in view of (3.18) to the equality:

$$\begin{pmatrix} 0 \\ v_2 - \ln y_1 \\ \vdots \\ v_n - \ln y_1 \end{pmatrix} = \underbrace{\begin{pmatrix} -u_1, & x_1 - 1 \\ & \vdots \\ -u_n, & x_n - 1 \end{pmatrix}}_{\boldsymbol{F}} \begin{pmatrix} b \\ c \end{pmatrix} + \boldsymbol{\varepsilon},$$

where vector $\boldsymbol{\varepsilon}$ and the variables y_i, u_i and v_i have the same meaning as above.

The information matrix takes, according to (3.20) (when taking the weights $w_i = 1$), the form

$$\boldsymbol{M}(\delta) = \sum_{i=1}^{n} \frac{1}{n} \begin{pmatrix} -u_i \\ x_i - 1 \end{pmatrix} (-u_i, x_i - 1) = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} u_i^2, & -u_i(x_i - 1) \\ -u_i(x_i - 1), & (x_i - 1)^2 \end{pmatrix}$$
$$= \frac{1}{n} \begin{pmatrix} \sum_{i=1}^{n} u_i^2, & -\sum_{i=1}^{n} u_i(x_i - 1) \\ -\sum_{i=1}^{n} u_i(x_i - 1), & \sum_{i=1}^{n} (x_i - 1)^2 \end{pmatrix},$$

from which (cf. (3.19), when taking $\frac{\sigma^2}{N} = 1$)

$$\mathrm{Var}_{III}(\hat{\boldsymbol{\Theta}}(\boldsymbol{Y}_\delta)) \doteq 1.0 \cdot 10^{-5} \begin{pmatrix} 0.868651, & 0.001135 \\ 0.001135, & 0.000001 \end{pmatrix},$$

and subsequently (see (3.11), (3.12))

$$\mu(b_{III}) \doteq \sqrt{0.868651} \doteq 0.002947, \tag{3.26}$$

$$\mu(c_{III}) \doteq \sqrt{0.000001} \doteq 0.000004. \tag{3.27}$$

ad IV) (three free parameters A, b, c)

Logarithmizing the equation (3.2), we arrive in view of (3.18) to the equality:

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \underbrace{\begin{pmatrix} 1, & -u_1, & x_1 \\ & \vdots & \\ 1, & -u_n, & x_n \end{pmatrix}}_{F} \begin{pmatrix} a \\ b \\ c \end{pmatrix} + \boldsymbol{\varepsilon}.$$

The vector $\boldsymbol{\varepsilon}$ and the variables y_i, u_i and v_i have the same meaning as above.

The information matrix takes, according to (3.20) (when taking the weights $w_i = 1$), the form

$$\boldsymbol{M}(\delta) = \sum_{i=1}^{n} \frac{1}{n} \begin{pmatrix} 1 \\ -u_i \\ x_i \end{pmatrix} (1, \ -u_i, \ x_i - 1) = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1, & -u_i, & x_i \\ -u_i, & u_i^2, & -u_i x_i \\ x_i, & -u_i x_i, & x_i^2 \end{pmatrix}$$

$$= \frac{1}{n} \begin{pmatrix} n, & -\sum_{i=1}^{n} u_i, & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} u_i^2, & \sum_{i=1}^{n} u_i^2, & -\sum_{i=1}^{n} u_i x_i \\ \sum_{i=1}^{n} x_i, & -\sum_{i=1}^{n} u_i x_i, & \sum_{i=1}^{n} x_i^2 \end{pmatrix},$$

from which (cf. (3.19), when taking $\frac{\sigma^2}{N} = 1$)

$$\mathrm{Var}_{IV}(\hat{\boldsymbol{\Theta}}(\boldsymbol{Y}_\delta)) \doteq \begin{pmatrix} 0.018689, & 0.002724, & 0.75 \cdot 10^{-8} \\ 0.002724, & 0.000405, & 0.12 \cdot 10^{-8} \\ 0.75 \cdot 10^{-8}, & 0.12 \cdot 10^{-8}, & 0.48 \cdot 10^{-12} \end{pmatrix},$$

and subsequently (see (3.14)–(3.16))

$$\mu(\ln A_{IV}) \doteq \sqrt{0.018689} \doteq 0.136710, \tag{3.28}$$

i.e. (cf. Kubáček, Kubáčková, 2000, p. 215)

$$\mu(A_{IV}) = \frac{\partial e^a}{\partial a} \Big|_{a = \ln A_{IV}} \mu(\ln A_{IV}) = A_{IV} \mu(\ln A_{IV})$$

$$\doteq 2.2939 \cdot 0.13671 \doteq 0.313600, \tag{3.29}$$

$$\mu(b_{IV}) \doteq \sqrt{0.000405} \doteq 0.020148, \tag{3.30}$$

$$\mu(c_{IV}) \doteq \sqrt{0.48 \cdot 10^{-12}} \doteq 0.000007. \tag{3.31}$$

Summing up the above calculations, we can give the following second conclusion.

Conclusion 3.2. *The standard deviation* μ *of parameters a, b, c in formulas* (3.2)–(3.5) *of MAL takes (by means of* (3.19)–(3.21)*) the respective forms* (3.22) *or* (3.23)*,* (3.25) *or* (3.26)*,* (3.27) *or* (3.28)*,* (3.30)*,* (3.31)*. The deviations are universal in the sense that they are independent of concrete data* $x_1, x_2, \ldots; y_1, y_2, \ldots;$ $z_1, z_2, \ldots,$ *and subsequently of concrete values of parameters* b_I *or* A_{II}*,* b_{II} *or* b_{III}*,* c_{III} *or* A_{IV}*,* b_{IV}*,* c_{IV}*, respectively. The deviations* $\mu(A_{II})$ *and* $\mu(A_{IV})$ *are no longer universal, but they satisfy the respective estimates* (3.24) *and* (3.29)*).*

Remark 3.2. There is a question about the correct usage of regression models I) and III) in the above mentioned form. It concerns the fixed value y_1 which can be further regarded as a constant by means of which the remaining unknown parameters are estimated. Nevertheless, since the value y_1 represents the realization of a random variable, it contains the measurement error which is subsequently transferred to the regression model. Unfortunately, the regression model does not take this "double error" into account. By this reason, we must consider all the resulting values obtained by means of formulas (3.4) and (3.5) only as those conditioned by the value of y_1.

3.4   Concluding remarks

In Sections 3.2 and 3.3, the comprehensive analysis was done which provides a comparison of four formulas (3.2)–(3.5) from the point view of the best approximation of data as well as of the accuracy of parameters A, b, c estimates. These two approaches are based on the least square method, but provide somewhat different goals. Thus, we provide an information about a suitability of the usage of models which are evaluated from two perspectives.

As already pointed out, from the point of view of the first perspective, it might be expected that the best results can be obtained by means of formula (3.2), while the worst by means of formula (3.5). In view of the analysis in Section 3.3, the situation is however much more delicate.

Peter Angermann, *A. I. (Nohandsfractal)*, 1989, oil on canvas

4 LANGUAGE FRACTALS

4.1 Historical stimulation

Postulating the self-similarity property as the definition of a fractal, Luděk Hřebíček conjectured a *weak version of a fractal structure of language* in the sense that formally the same Menzerath–Altmann law holds on every linguistic level (see e.g. Hřebíček, 2002, and the references therein).

Furthermore, he noticed that the reciprocal value of the shape parameter at the truncated formula of the Menzerath–Altmann law can be expressed as a self-similarity dimension by means of the Moran–Hutchinson formula (see Hřebíček, 1994, 1998b).

The same relationship was already detected by Benoît B. Mandelbrot for the exponent at the Zipf–Mandelbrot law (see Mandelbrot, 1983, Chapter XI; Mandelbrot, 2000, Chapter 12, and the references therein).

On the other hand, in view of these observations, the interpretation of both linguistic laws in fractal terms was only possible for structures exhibiting the very restrictive exact (strict) self-similarity property. More realistic interpretation was proposed by ourselves in Andres (2009) and Andres, Rypka (2012), for linguistic structures with a more liberal cyclic self-similarity.

In this chapter, the above discoveries will be recalled in a more detail at first. Then they will be critically discussed in a new light. Our goal is not just a description of a language structure under consideration, but also an attempt for a possible explanation of the meaning of the computed fractal dimensions to given texts. To characterize author's style or to distinguish between originals and falsa or so would be, however, too ambitious. We understand that after the first steps in this field, based especially on linguistic experiments (see e.g. Hřebíček, 2002, 2007; Eftekhari, 2006; Köhler, 2008; Andres, Benešová, Kubáček, Vrbková, 2012; Andres, Benešová, 2011, 2012), it is still a long way to go.

4.2 Menzerath–Altmann law: complete formula vs. truncated formula (continued discussion from Chapter 3)

The Menzerath–Altmann law (MAL) expresses the relationship between the length x of a *construct* (i.e. a unit on an upper linguistic level) and the length y of its *con-*

stituents (i.e. units on a lower linguistic level). The *complete formula* of MAL takes the form (cf. Altmann, 1980; Wimmer, Altmann, Hřebíček, Ondrejovič, Wimmerová, 2003)

$$y = Ax^{-b}e^{cx},$$

or, equivalently (for $b \neq 0$),

$$\frac{1}{b} = \frac{\log x}{\log\left(\frac{A}{y}\exp(cx)\right)} = \frac{\ln x}{\ln\left(\frac{A}{y}\exp(cx)\right)},$$

where A, b, c are real parameters, $e = 2.718\ldots$ is the Euler number.

For $c = 0$, it reduces to the *truncated formula*, i.e.

$$y = Ax^{-b}, \quad \text{resp. (for } b \neq 0), \ \frac{1}{b} = \frac{\log x}{\log \frac{A}{y}}.$$

In the particular case, when $A = y(1) = y_1$, we get the simplest formula

$$y = y_1 x^{-b}, \quad \text{resp. (for } b \neq 0), \ \frac{1}{b} = \frac{\log x}{\log \frac{y_1}{y}},$$

with only one free parameter b, usually called a *shape parameter*.

As demonstrated in the foregoing chapter, although the application of the complete formula is certainly better from the point of view of the optimal approximation of given data, it is quite opposite from the point of view of the accuracy of parameter b (which is crucial for our investigation) estimates. Thus, a suitability of the usage of the above formulas must be evaluated together from two perspectives at least. In our linguistic experiment in Andres, Kubáček, Machalová, Tučková (2012), rather surprisingly, the simplest formula occurs to be optimal among all from the statistical point of view. We believe that this particular case can hold in general, provided the structure of given data is statistically significant.

Moreover, since the shape of the graph of the function $y = Ax^{-b}e^{cx}$ with $c > 0$ does not often reflect the verbal form of MAL, for larger values of x (cf. Andres, Kubáček, Machalová, Tučková, 2012), it can be only used for small integers x. On the other hand, the "optimal" usage of the simplest formula brings a serious obstruction, for the related regression model. Namely, since the value y_1 represents the realization of a random variable, it contains the measurement error which is automatically transferred into the regression model. Subsequently, all the resulting values must be only considered as those conditioned by y_1.

The second criticism about the usage of the complete formula concerns its derivation. It seems to us that the implementation of an additional parameter c via the exponential function e^{cx} is rather artificial (i.e. without any satisfactory linguistic explanation), just because of a simple application of a linear regression technique. For more details concerning this technique (see e.g. Montgomery, Peck, Vinig, 2006).

Perhaps more appropriately, instead of the initial equation

$$\frac{\dot{y}}{y} = \frac{-b}{x} + c,$$

one should rather start from the one of the form

$$\frac{\dot{y}}{y} = \frac{-b}{x+c},$$

because in this way parameter c can be interpreted as something like a "mental overrunning" in our speach yielding the construct completion. Thus, the related formula would take the form

$$y = A(x+c)^{-b},$$

or, equivalently (for $b \neq 0$),

$$\frac{1}{b} = \frac{\log(x+c)}{\log \frac{A}{y}}.$$

Unfortunately, a direct application of the linear regression technique fails here, because after its logarithmization, the equation becomes nonlinear. On the other hand, the obtained nonlinear equation can be still linearized, and subsequently the linear regression technique can be applied to a linearized equation.

4.3   Zipf–Mandelbrot law: isomorphism of form

The Zipf–Mandelbrot law (ZML) describes the distribution of word frequencies for regular lexicographic trees. It takes the form (see e.g. Mandelbrot, 1983, Chapter XI; Mandelbrot, 2000, Chapter 12; Manin, 2009; Montemurro, 2001, 2004; Wimmer, Altmann, 2005, and the references therein)

$$U = P(\rho + V)^{-\frac{1}{D}}$$

or, equivalently,

$$D = \frac{\log(\rho + V)}{\log \frac{P}{U}},$$

where ρ denotes the order of a word (words were ordered in a decreasing way according to their frequencies) with probability U and P, V, D are real constants. The *regularity of trees* means that each branching is related to a single word and the probability weight on the k-th level takes the form $U = U_0 r^k$, where $0 < r < 1$, and U_0 is such that the sum of all probability weights is equal to 1.

Observe that, despite their meaning, the last two formulas, i.e. $y = A(x+c)^{-b}$ and $U = P(\rho + V)^{-1/D}$, are formally the same. Thus, ZML can be again obtained as a general solution of the linear ordinary differential equation

$$\frac{\dot{U}}{U} = \frac{-1/D}{\rho + V}, \quad \text{where } \dot{U} = \frac{\mathrm{d}V}{\mathrm{d}\rho}.$$

As already pointed out in Chapter 2, the effect of MAL can be detected not only in various domains of linguistics, but also in nonlinear biology, sociology, psychology, etc. The same is true for ZML, as explained in economic terms in Chapter 13 in Mandelbrot (1983) and Chapter 12 in Mandelbrot (2000), for a mathematically equivalent form of *Pareto's law*, i.e.

$$\rho = -V + U^{-D}P^{D}.$$

Such an isomorphism documents the well known principle of a least effort *reflecting the economization of nature* (see Wimmer, Altmann, 2005, and the references therein) or a certain sort of a *conservation law*.

There is, however, still one more remarkable isomorphism to the *Moran–Hutchinson formula* (for more details, see e.g. Barnsley, 2013, Chapter V), for the computation of a *self-similarity dimension* of a fractal, namely

$$D = \frac{\log m}{\log \frac{1}{r}},$$

where m denotes the number of parts on each scale (number at contractions of the generating iterated function system) and r stands for the length of each part (contraction factor), provided the *open set condition* is satisfied (i.e. if the fractal set is either totally disconnected or its parts are at least just touching). Moreover, because of the above interpretation, and subsequently a possible visualization by means of iterated function systems, $(x + c)$ resp. $(\rho + V)$ should be a positive integer.

This isomorphism was observed for the first time by Mandelbrot (see Mandelbrot, 1983, Chapter XI; Mandelbrot, 2000, Chapter 12, and the references therein) w.r.t. ZML and, independently, by Hřebíček (see Hřebíček, 1994) w.r.t. MAL.

Despite the evident correspondences

$$m \sim \rho + V, \quad m \sim \frac{U}{P} = (\rho + V)^{-\frac{1}{D}}$$

or

$$D \sim \frac{1}{b}, \quad m \sim x + c, \quad r \sim \frac{y}{A} = (x + c)^{-b}$$

(or, for the complete formula of MAL, $D \sim \frac{1}{b}$, $m \sim x$, $r \sim \frac{y}{A \exp(cx)} = x^{-b}$), the application of fractal geometry to ZML or MAL might be only theoretical, because the exact self-similarity would require nonrealistic regular trees or the same values of parameters A, b, c on all linguistic levels, respectively.

4.4 Fractal analysis of texts: cyclic self-similarity hypothesis

Nevertheless, assume that on n scaling levels $i = 1, 2, \ldots, n$ we have

$$U_i = P_i (\rho_i + V_i)^{-\frac{1}{D_i}} \quad \text{or} \quad y_i = A_i (x_i + c_i)^{-b_i}$$

(or, for the complete formula of MAL, $y_i = A_i x_i^{-b_i} e^{c_i x_i}$) and that (cyclicity hypothesis) $U_i = U_{i+rn}$, $P_i = P_{i+rn}$, $\rho_i = \rho_{i+rn}$, $V_i = V_{i+rn}$, $D_i = D_{i+rn}$, for $r = 1, 2, \ldots$, or $y_i = y_{i+rn}$, $A_i = A_{i+rn}$, $x_i = x_{i+rn}$, $c_i = c_{i+rn}$, $b_i = b_{i+rn}$, for $r = 1, 2, \ldots$, i.e. that the situation is cyclically repeated in the blocks of n-levels (which is a purely mathematical but more realistic assumption). In this way, we arrive at the following interpretation of ZML or MAL in fractal terms (in case of MAL, see Andres, 2009; Andres, Rypka, 2012):

ZML:

$$D := \frac{n}{\frac{1}{D_1} + \frac{1}{D_2} + \cdots + \frac{1}{D_n}},$$

$m := (\rho + V)^{kn}$, provided $\rho + V = \rho_1 + V_1 = \rho_2 + V_2 = \cdots = \rho_n + V_n$ are such that $(\rho + V)$ is a positive integer,

$$r = r_1 \ldots r_n := \prod_{i=1}^{n} \left(\frac{U_i}{P_i} \right)^k = (\rho + V)^{-k \sum_{i=1}^{n} \frac{1}{D_i}},$$

where $k \geq \max_{i=1,2,\ldots,n} D_i$ is a suitable positive integer,

$$D := \frac{n \log(\rho + V)}{\sum_{i=1}^{n} \log \frac{P_i}{U_i}}, \quad \text{resp.} \quad \rho = -V + \exp\left(\frac{D}{n} \sum_{i=1}^{n} \ln \frac{P_i}{U_i} \right),$$

which is a generalized *Mandelbrot's* resp. *Pareto's formula*, because for $P = P_1 = P_2 = \cdots = P_n$ and $U = U_1 = U_2 = \cdots = U_n$, we get

$$D = \frac{\log(\rho + V)}{\log \frac{P}{U}}, \quad \text{resp.} \quad \rho = -V + U^{-D} P^D.$$

Analogously,

$$D := \frac{n}{b_1 + \cdots + b_n} = \frac{n \log(x + c)}{\sum_{i=1}^{n} \log \frac{A_i}{y_i}},$$

$m := (x + c)^{kn}$, provided $x + c = x_1 + c_1 = x_2 + c_2 = \cdots = x_n + c_n$ are such that $(x + c)$ is a positive integer,

$$r = r_1 \ldots r_n := \prod_{i=1}^{n} \left(\frac{y_i}{A_i} \right)^k = (x + c)^{-k \sum_{i=1}^{n} b_i},$$

where $k \geq \max_{i=1,2,\ldots,n} \frac{1}{b_i}$ is again a suitable positive integer.

MAL:

$$D := \frac{n}{b_1 + \cdots + b_n} = \frac{n \log x}{\sum_{i=1}^{n} \log \left(\frac{A_i}{y_i} \exp(c_i x) \right)},$$

$m := x^{kn}$, provided $x = x_1 = x_2 = \cdots = x_n$,

$$r = r_1 \ldots r_n := \prod_{i=1}^{n} \left(\frac{y_i}{A_i \exp(c_i x)} \right)^k = x^{-k \sum_{i=1}^{n} b_i},$$

where $k \geq \max\limits_{i=1,2,\dots,n} \frac{1}{b_i}$ is a suitable (sufficiently large) positive integer in order the open set condition to be satisfied.

Roughly speaking, if $1/D_i > 0$ resp. $b_i > 0$, for every $i = 1,2,\dots,n$, then we can speak about the given language structures as *language fractals*, because their models represent approximations of the particular associated mathematical fractals whose self-similarity dimension D is described above. For more details, see Andres (2009) and Andres, Rypka (2012).

If, reversely, $1/D_i \leq 0$ resp. $b_i \leq 0$ holds, for some i, then either ZML resp. MAL fails on the i-th level or there is missing an intermediate linguistic level which should have been taken for appropriate constituents, instead of former ones (cf. Andres, 2010). Thus, the correct application of ZML resp. MAL can be tested (even for discovering possible new linguistic levels) by means of the sign of $1/D_i$ resp. b_i, $i = 1,2,\dots,n$. By new linguistic levels we mean, for instance, suprasentence levels which might be possibly higher than semantic constructs detected by Hřebíček (see Hřebíček, 2002, 2007; Wimmer, Altmann, Hřebíček, Ondrejovič, Wimmerová, 2003, and the references therein).

4.5 Computation of fractal dimension of structures: universality vs. specificity

For the simplest formula of MAL, when $A_i = y_i(1) = y_{1i}$, $i = 1,2,\dots,n$, and $c = c_1 = c_2 = \cdots = c_n = 0$, the self-similarity dimension D of the associated mathematical fractals can be simply estimated in terms of the length of given constructs xji and constituents y_{ji}, $i = 1,2,\dots,n$; $j = 1,2,\dots,p_i$, as follows (cf. Chapter 3):

$$D := \frac{n}{\sum_{i=1}^{n} b_i},$$

where

$$b_i \approx \frac{\ln y_{1i} \sum_{j=1}^{p_i} \ln x_{ji} - \sum_{j=1}^{p_i} \ln x_{ji} \ln y_{ji}}{\sum_{j=1}^{p_i} (\ln x_{ji})^2}, \quad i = 1,2,\dots,n.$$

Incorporating the weights

$$w_{ji} \approx \frac{z_{ji}}{\sum_{j=1}^{p_i} z_{ji}}, \quad i = 1,2,\dots,n,$$

corresponding to the j-th relative frequency $z_{ji} / \sum_{j=1}^{p_i} z_{ji}$, $j = 1,2,\dots,p_i$, into the approximative formula, we obtain

$$b_i \approx \frac{\ln y_{1i} \sum_{j=1}^{p_i} w_{ji} \ln x_{ji} - \sum_{j=1}^{p_i} w_{ji} \ln x_{ji} \ln y_{ji}}{\sum_{j=1}^{p_i} w_{ji} (\ln x_{ji})^2}, \quad i = 1,2,\dots,n.$$

The same is formally true for ZML in terms of trees, provided $P_i = U_i(1) = U_{1i}$, $i = 1,2,\dots,n$, and $V = V_1 = V_2 = \cdots = V_n = 0$. For the complete formula of MAL, we can obtain more complicated, but still explicit formulas estimating D (see

Chapter 3). On the other hand, for ZML in general, only numerical calculations can be made, because the linear regression technique does not apply there.

In linguistic experiments in spoken languages, the calculated values of D for novels and newspaper texts, or so, are higher than those for poems. Introductory chapters possess higher values of D than consequitive chapters, etc. On the basis of such empiric arguments which we believe to be so mainly due to semantics, we decided to call D as the *measure or degree of semanticity* of a given text, provided the lowest linguistic level is "reasonable" from the semantic point of view like the length of words calculated in the average number of syllables.

Mandelbrot in Mandelbrot (2000, Chapter 12) discusses the meaning of two possibilities: $D \geq 1$, when the text (in his case, the number of words) is finite, and $D < 1$, when it is infinite (which is practically nonrealistic). In order to simplify the calculations, we put in this section $V = V_1 = V_2 = \cdots = V_n = 0$. According to Mandelbrot, this means that the hierarchy trees are symmetric. Otherwise, i.e. in the asymmetric cases, he explains the role of V_i, $i = 1, 2, \ldots, n$, in terms of fractal lacunarity.

Eftekhari's fractal analysis of Shakespeare's works (see Eftekhari, 2006) consists of computing the fractal dimension D_F (always less than 1) and the Zipf dimension D_Z (always less than 2). In our analysis (again in spoken languages), the values of D are typically in one or even two orders higher. The reason of this discrepancy is simply the fact that Eftekhari computes different fractal dimensions, where the open set condition need not be satisfied, and so takes into account different relations. In a similar way, for a *box-counting dimension*, i.e. when refining the grid scales, for instance, the fractality of architectural objects can be calculated in order to clarify their aesthetic preference (cf. Lorenz, Andres, Franck, 2017, and the references therein). The images with a higher dimension were considered complex, while those with a lower dimension uninteresting. The preferable box-counting dimension of objects, whose supporting space is a two-dimensional plane, was found around 1.35 in den Heijer, Eiben (2010), 1.52 ± 0.23 in Draves, Abraham, Viotti, Abraham, Sprott (2008) which is in agreement with measurements in quoted papers there (1.51 ± 0.43, around 1.54, etc.).

It is a question whether to consider the lengths of constructs in the mean but absolute number of constituents (as we always did) or to "measure" the objects on concrete linguistic levels by means of suitable physical units, as proposed by Köhler in Köhler (1997) (cf. also Leopold, 2001 and Andres, 2010).

If, for instance, the length of semantic constructs would be endowed with kilo-something, the length of clauses with something, the length of words with centisomething and the length of syllables in milisomething, then after rescaling to the same unit (e.g. milisomething), the dimensions would dramatically change. Perhaps the most appropriate physical units in quantitative linguistics might be with this respect suitable (normalized) time units. Nevertheless, observe that in all the above formulas the average number of constituents was already normalized by means of $A_i(= y_{1i})$ resp. $P_i(= U_{1i})$.

4.6   Distance between two language fractals

We know from our experiments (from both spoken and sign languages) that there is unfortunately too sensitive dependence of the values of D on the way of segmentation of a given text. The same is true for various translations of the same text.

Hence, distinguishing by prime the values of shape parameters b_i, $i = 1, 2, \ldots, n$, of another text under consideration (like its translation) or the same text but segmented in a different way, it is useful not only to compare the related degrees of semanticity D but also to measure the *distance d between two language fractals* characterized by vectors $(b_1, b_2, \ldots, b_n)$ and $(b'_1, b'_2, \ldots, b'_n)$ as follows:

$$d\left(\left(\frac{1}{b_1}, \frac{1}{b_2}, \ldots, \frac{1}{b_n}\right), \left(\frac{1}{b'_1}, \frac{1}{b'_2}, \ldots, \frac{1}{b'_n}\right)\right) = \sqrt{\sum_{i=1}^{n} \left(\frac{1}{b_i} - \frac{1}{b'_i}\right)^2}.$$

In particular, for $n = 3$, we have

$$d\left(\left(\frac{1}{b_1}, \frac{1}{b_2}, \frac{1}{b_3}\right), \left(\frac{1}{b'_1}, \frac{1}{b'_2}, \frac{1}{b'_3}\right)\right)$$
$$= \sqrt{\left(\frac{1}{b_1} - \frac{1}{b'_1}\right)^2 + \left(\frac{1}{b_2} - \frac{1}{b'_2}\right)^2 + \left(\frac{1}{b_3} - \frac{1}{b'_3}\right)^2}.$$

4.7   Concluding remarks

Despite the above criticism, we understand that in quantitative linguistics something like the complete formula of MAL is necessary, because its reduced form does not often fit the given data on lower linguistic levels (from the words, resp. signs, downwards).

According to the personal experience of Prof. Reinhard Köhler with linguistic data on all levels, the power law form is more appropriate to higher linguistic levels (i.e. to those associated with more semantic units) whereas the pure exponential part of the formula corresponds to the "material" or "energetic" levels (e.g. when a sound duration is involved). The application of a complete formula is so, in view of his arguments, appropriate especially to intermediate levels.

On the other hand, the complete formula of MAL might be sometimes replaced by the following one:
$$y = A(x + c)^{-b}.$$

Roughly speaking, one should make a suitable choice between easy calculations, when a linear regression technique directly applies as a theory, and perhaps linguistically a more justified law, when all parameters must be calculated in a more complicated way. A compromise could be possible, provided the value of parameter c in the above formula can be detected, at least for some sorts of texts, empirically. This assumption seems to be, however, practically nonrealistic.

After all, the unified formula, involving four free parameters A, b, c_1, c_2, which is a general solution of the differential equation

$$\frac{\dot{y}}{y} = \frac{-b}{x+c_1} + c_2,$$

takes the form

$$y = A(x+c_1)^{-b}e^{c_2 x}.$$

Summing up briefly the above informations, for fractal analysis of texts (speeches), there is still a challenge to make a suitable choice among the possibilities concerning:

- the "optimal" formula of MAL (resp. ZML),
- whether or not to incorporate the weights into calculations,
- the appropriate way of segmentation of a given text.
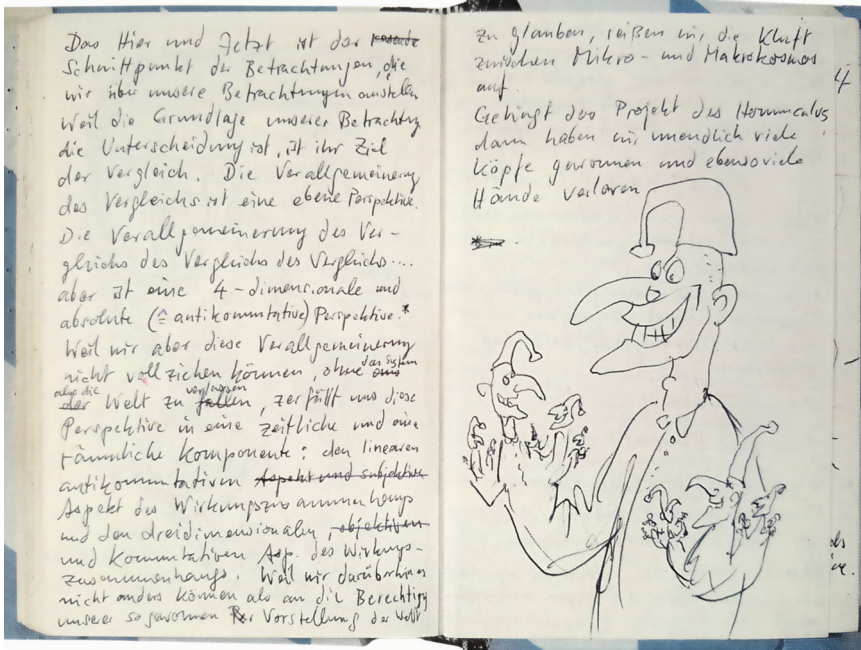
Peter Angermann, preparatory drawing of *A. I. (Nohandsfractal)* with author's comments, 1989

5 METHODOLOGY

5.1 Basic steps, especially visualization

As a methodological design for the fractal analysis of a sign language text (speech), it would be optimal to examine at least its three levels ($i = 1, 2, 3$) by means of the Menzerath–Altmann formula. However, unlike to spoken languages, in a sign language it seems to be delicate to detect a supra-sentence level on one side, and to identify constituents for signs on the other hand. For the latter case, we will come back at the end of this chapter.

Nevertheless, to reach the goal of our Experiment 1 (see Chapter 7 below), we need to find the parameters A_i, b_i, c_i, $i = 1, 2, 3$, for both the truncated and complete formulas of the Menzerath–Altmann law (MAL). Calculating and commenting them on merely two steps of the following algorithm which is also summarised in the flow chart (diagram).

Nine steps of the algorithm are as follows:

STEP 1. The choice of the sample text and explaining the choice.

STEP 2. Determination of the sample units and explaining it. Units have to be defined unambiguously; the notion of the unit has to be in accordance with common linguistic definitions, and if not, it has to be carefully justified; the determination of units has to be rigidly kept throughout the whole experiment; and each sample member has to be taken into account, yet not calculated twice.

STEP 3. Verifying the representativeness of the sample length.

STEP 4. Quantifying the text so that it is possible to extract the variables x_i and y_i for every $i = 1, 2, 3$ from it.

STEP 5. Calculating the parameters A_i, b_i, c_i, $i = 1, 2, 3$, for both the truncated and complete formulas of the Menzerath–Altmann law by means of the above described statistical and numerical methods (see Chapter 3).

STEP 6. Testing the model reliability by means of the statistical methods (see Chapter 6 below).

STEP 7. Interpreting the parameters A_i, b_i, c_i, $i = 1, 2, 3$ in the fractal analysis.

STEP 8. Visualizing language structures by means of approximating the mathematical fractals with a given dimension.

STEP 9. Interpreting the visualizations of language structures.

Before we proceed to the diagram, let us briefly explain Step 8 – visualization. For more details, see our paper Andres, Rypka (2012) and Andres, Benešová, Kubáček, Vrbková (2012).

In view of the desired fractal analysis, the collection

$$A_1 := F_1([0,1]), \ A_2 := F_2 \circ F_1([0,1]), \ A_3 := F_3 \circ F_2 \circ F_1([0,1]) = F([0,1]),$$

where

$$F_i(x) := \bigcup_j {}_i f_j(x), \ \ {}_i f_j \colon [0,1]^k \to [0,1]^k, \ \ k \ge \max\left\{\tfrac{1}{b_1}, \tfrac{1}{b_2}, \tfrac{1}{b_3}\right\},$$

$${}_i f_j(x) := rx + \frac{1}{x}j, \ \ r = x^{-k(b_1+b_2+b_3)},$$

$$j = (j_1, \ldots, j_k), \ \ j_l \in \{0, 1, \ldots, x-1\}, \ \ i = 1, 2, 3,$$

can be already regarded, under the correspondence described in Chapter 4, as the first step of a visualized structure of linguistic objects on $n = 3$ linguistic levels, characterized by the coefficients A_i, b_i, c_i ($i = 1, 2, 3$) at the MAL.

Observe that, for $A_{s3} := F^s([0,1])$, we have, according to the arguments above, that the limit of $d_H(A_{s3}, \boldsymbol{A})$ is for s tending to infinity, 0 and the above estimate, for the Hausdorff distance $d_H(A_{s3}, \boldsymbol{A})$ between A_{s3} and $\boldsymbol{A}$, holds.

Moreover, F^s consists of x^{3ks} contractions with the same factor

$$r := x^{-k(b_1+b_2+b_3)},$$

where k is a suitable positive constant (in order the open set condition to be satisfied).

For visualization of the above collection A_1, A_2, A_3 and the sets $A_{s3}, s = 1, 2, \ldots$, for the given initial set $[0,1]$, we make use of the very last iteration. The initial set does not affect the output attractor, yet can be consequential for plotting iterations. For simplification it is advantageous to determine simple sets with a few points. For instance, line segments, which are defined with two points, can be used. By substituting into the formulas, we can calculate the coordinates of the points (counter images), whose number is x^k times as much. In the s-th step, we get $2x^{3ks}$ points. We are able to calculate in this way only a few iterations, but usually in a few steps succeeding iterates are indistinguishable. The length of the line segments in the s-th step is $\prod_{k=1}^{3s} r_{i_k} i_k \in \{1, 2, 3\}$.

When we get the pairs of the particular points, we can easily plot the line segments being the last iterates out of them. Because of the monitor and eye resolution, to perform contractions in the line segments shorter than thousandths of the plotted interval length is no use.

In our case, we would like to consider the composition $F = F_3 \circ F_2 \circ F_1$ of three Hutchinson–Barnsley maps and its projection into two-dimensional space, i.e. we take x^2 similitudes. Creating one system by composing $n = 3$ maps F_1, F_2, F_3 would, needless to say, be feasible, and would contain x^6 mappings; nonetheless, the possibility of modelling the segmentation of language structures would be lost. Any composition of contractions (similitudes) is again a contraction (similitude),

i.e. there is an attractor of the composition F, and the iterations of a line segment initial set will be composed of line segments. Thus, we create the sequence

$$[0,1], \; F_1([0,1]), \; F_2(F_1([0,1])), \; F_3(F_2(F_1([0,1]))) = F([0,1]), \; F_1(F([0,1])),$$
$$F_2(F_1(F([0,1]))), \; F^2([0,1]) = F(F([0,1])), \; F_1(F^2([0,1])), \; \ldots$$

But we plot solely the iterates of the composed mapping $F^s([0,1])$.

One can easily draw iterates of line segments in MATLAB. As already pointed out, it only suffices to the ends of line segments by mappings determined by F_1, F_2, F_3, because the MATLAB instruction line connects the ends.

Two first approximations of a visualized (sign) language fractal can look as those depicted in Figures 5.1 and 5.2. Its fractal dimension is $D = 9.465$.
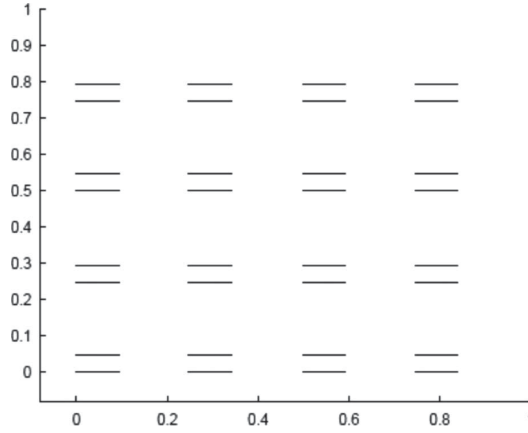
Figure 5.1: Two-dimensional projection of the first approximation of a (sign) language fractal
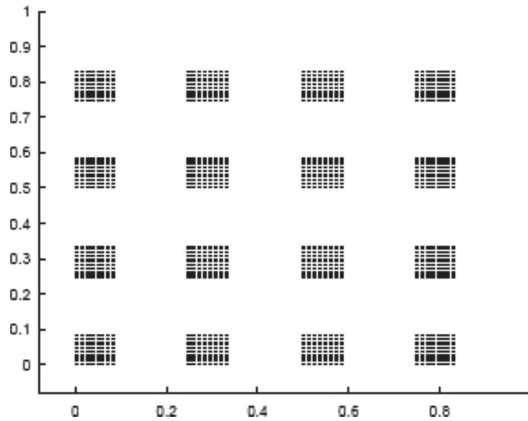
Figure 5.2: Two-dimensional projection of the second approximation of a (sign) language fractal
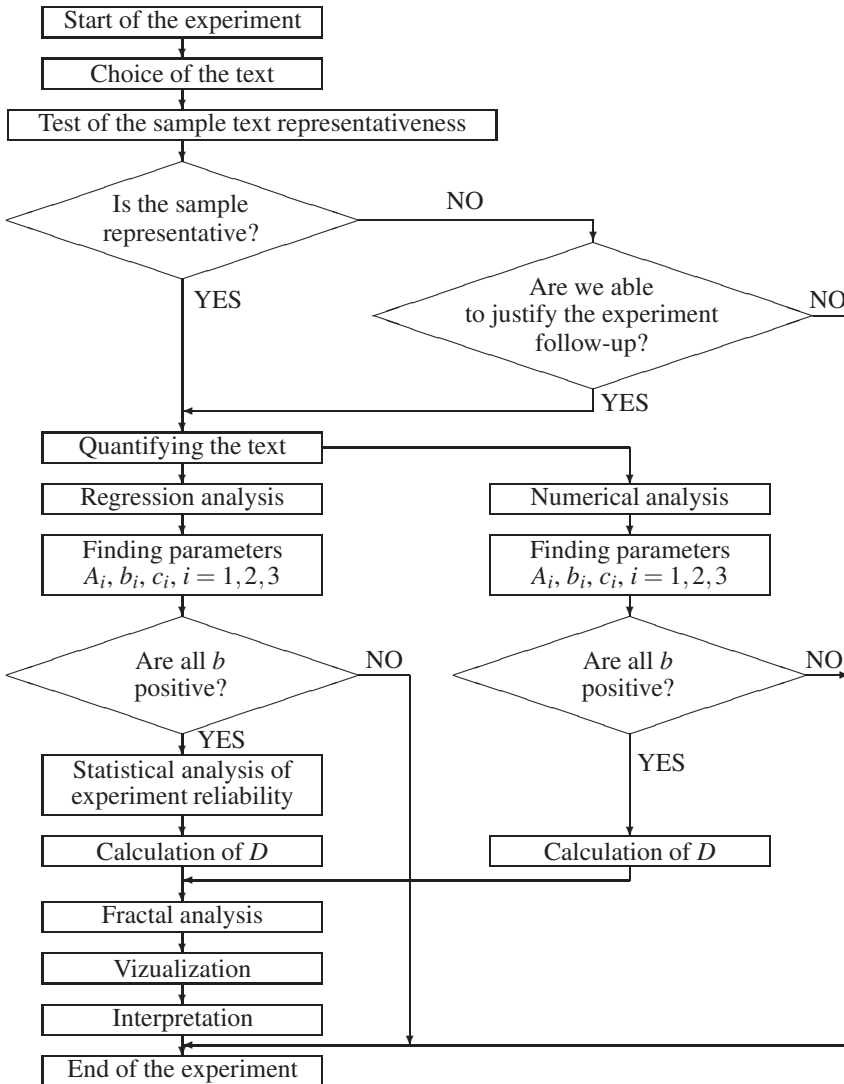
Figure 5.3: The flow chart depicting the steps of the fractal analysis of the text

5.2 Looking for the constituents of signs

Due to the simultaneous character of the manual signs (as mentioned above), one of the most challenging tasks is to develop a method of counting the length of signs (the number of constituents). One of the proposed methods is a simple sum of all morphemes which are produced over its course, according to these sign parameters:

- sign location in the articulation space (TAB)
- shape of the articulating hand/hands (DEZ)

- orientation of the palm/palms (ORI1)
- orientation of fingers (ORI2)
- movements of the hand/hands (SIG)
- mutual hand arrangement in two-handed signs (HA)

Table 5.1: The design of the analysis and notation of the total morpheme number in the one-handed sign (CHOOSE)

	Number of morphemes[*]	
Parameter	Right hand	Left hand
TAB	1	0
DEZ	2	0
ORI1	1	0
ORI2	1	0
SIG	1	0
HA	0	
Total	6	

CHOOSE

[*]The number of morphemes in the above-mentioned examples indicates the number of values which can be obtained by a certain parameter in a sign. The place of articulation of the sign CHOOSE is, for example, unchanged (TAB = 1) as well as the palm orientation (ORI1 = 1) and the orientation of the fingers (ORI2 = 1) while the movement of a hand is not repeated (SIG = 1). The original shape of the hands, however, is changed (DEZ = 2).

A number of studies performed on ASL indicate that certain hand shapes can affect the perception of sign complexity (cf. Brentari, 2011; Brentari et al., 2017). In the case of the Czech sign language, this has not yet been proven. The table presented above indicates that the total number of morphemes is significantly influenced by the number of hands (which are used to create a sign) and the number of phonologically important phonemes in the individual parameters of signs (morphemes). The question remains, however, as to whether the two-handed sign (with a higher total sum of morphemes) is actually perceived by users of sign language as significantly more difficult to learn and remember than the one-handed sign with a logically lower number of morphemes.

Figure 5.3: The dependence of the average value of the complexity of the sign on the number of morphemes

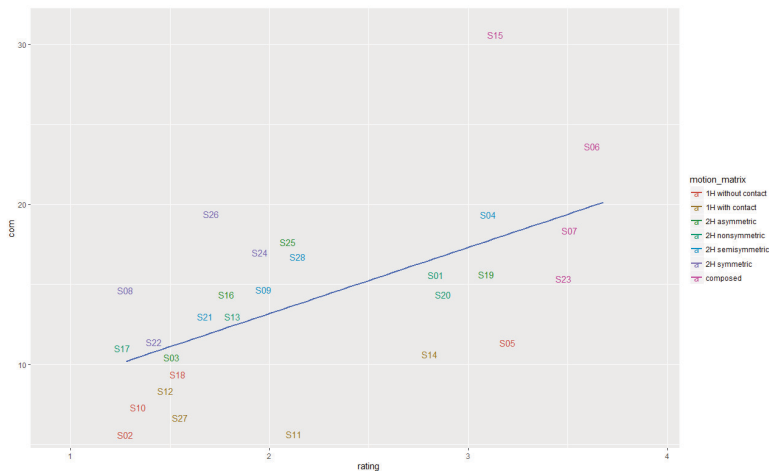We consequently conducted a study aimed at detecting the dependence between the number of sign morphemes, their belonging to defined motion matrices, and the assessment of their complexity by the users (see Chapter 10 below and cf. Langer, Rypka, 2017). The linear dependency of the complexity of the sign on the number of its morphemes (see Figure 8 below) makes obvious, although not entirely essential (the correlation reaches the value 0.4), the distribution of the values of the average assessment of the complexity of the signs. In addition, the non-standard placement of signs S08, S09, S24, S25, S26 and S28 is apparent, where they are assessed as relatively simple, despite a rather high number of morphemes. Due to the fact that they are two-handed symmetrical signs (2H symmetrical), the question emerges if there is an unwanted overrating of the total number of morphemes among these signs, since it is basically a mirror image and motion with both hands.

The performed research confirmed the original theoretical assumption that the number of morphemes (phonemes which carry meaning) in Czech sign language signs proportionally influences the subjective perception of the complexity of the presented signs. The number of morphemes was stated at the beginning of the research in a way that referred to the sum of phonologically important phonemes occurring during the articulation of a sign within the individual parameters of a sign for both the dominant and the non-dominant hand. An important factor in perceiving the complexity of the signs may be specific hand shapes or motions, but our study does not currently address this phenomenon.

The results stated above were also essential for the execution of the quantitatively linguistic analysis of Czech sign language, for which the number of morphemes of the individual signs is the lowest level of the observed hierarchical structure.

Our experiments thus far have unfortunately only shown that the relationship between the length of the constituents (counted in the number of morphemes in the signs) and the length of the constructs (counted in the number of signs) do not match the MAL. Our next attempt will therefore be further experimental segmentation and subsequent identification of the number of constituents below the sign's level.

To detect supra-sentence levels of a sign language is another challenge for our future research.

Peter Angermann, *Interaction*, 1990, animated painting

6 STATISTICAL VERIFICATION OF RESULTS

6.1 Convention for the usage of the MAL formulas

In our former Chapter 3, the following four formulas of the *Menzerath–Altmann law* (one of the generally accepted linguistic laws formulated in a quantitative way):

   I) $y = y_1 x^{-b}$,

   II) $y = Ax^{-b}$,

   III) $y = y_1 x^{-b} e^{c(x-1)}$,

   IV) $y = Ax^{-b} e^{cx}$,

where A, b, c are real parameters, were examined from two perspectives. The first goal was related to the best approximation of given data, while the second one was especially concentrated on the accuracy of a calculated shape parameter b which is necessary for the fractal analysis of the text.

For the conclusions dealing with optimal strategies (i.e. the balance between rigorousness and simplicity), we have always presumed that the structure of given data is statistically significant. In particular, in Section 3.3 titled "Comparison of accuracy of parameter estimations", the term $\frac{\sigma^2}{N}$ in the formula

$$\mathrm{Var}(\widehat{\mathbf{\Theta}}(\mathbf{Y}_\delta), N) := \frac{\sigma^2}{N} \mathbf{M}^{-1}(\delta)$$

for covariance matrix of the regression parameter estimates, was assumed to be equal to 1.

In the present chapter, only formulas I), II) and IV) will be taken into account. On the other hand, the length y of constituents will be considered not only as the mean value $\bar{y}$ as in Chapter 3, but also as the set $\{y\}$ of partially averaged values of y, which we call *semi-averaging*. More concretely, for each construct, we make individually the averaging of its associated constituents. For instance, for the sentence constituted of 2 clauses with lengths 2 and 3 words (signs), we have $x = 2$ and $y = (2+3)/2 = 2.5$. In this way, we make a certain normalization, because the frequencies of constructs will be always equal to 1. On the other hand, in the case without any averaging, the lengths of constituents would be integers, but the frequencies of constructs should be taken, rather curiously, noninteger-

valued, in general. Therefore, the formulas I), II) and IV) as above will concern this time the situation with semi-averaging, while the "bar" formulas, i.e.

$\overline{\text{I}}$) $\bar{y} = \bar{y}_1 x^{-b}$,

$\overline{\text{II}}$) $\bar{y} = Ax^{-b}$,

$\overline{\text{IV}}$) $\bar{y} = Ax^{-b}e^{cx}$,

will be those with the averaged values $\bar{y}$ of $y \in \{y\}$, i.e. the averaged value of semi-averaged values.

All formulas are again supposed to be applied to three pairs of linguistic levels. For instance, for spoken languages: level 1, i.e. semantic constructs[1] vs. clauses, level 2, i.e. clauses vs. words, level 3, i.e. words vs. syllables. For a sign language, we would also like to examine at least three pairs of linguistic levels like: sentences vs. clauses, clauses vs. signs, signs vs. morphemes, but as already pointed out (see e.g. the arguments in the foregoing Chapter 5), the multidimensional (simultaneous) character of the lowest level brings some serious obstructions.

In this chapter, the structure of given data will be mainly examined in detail from the statistical point of view. More concretely, their normality will be tested by the Kolmogorov–Smirnov test (see e.g. Thode, 2002), their homoscedasticity resp. heteroscedasticity will be tested by the White test (see e.g. White, 1980). The quality of fitting will be checked by means of the residual standard error, the root mean square error, the normalized root mean square error and the coefficient of determination (see e.g. Montgomery, Peck, Vinig, 2006).

The appropriate model for the calculation of real parameters A, b, c should be selected just on the basis of such an analysis. In the case of non-normality, the related confidence intervals should be calculated by the bootstrap technique (for more details, see e.g. Efron, Tibshirani, 1993).

6.2   Some preliminaries from statistics

In the entire chapter, the symbol N denotes the total number of observed values $y_j \in \{y\}$, while n stands for the number of different construct lengths. Let us note that N and n can differ for different linguistic levels.

Using the logarithmic transformation, all models under consideration can be linearized. Thus, for the models I), II) and IV), when the values of constituent lengths are semi-averaged, we obtain linear models of the form:

ad I) $\ln Y_j = \ln \bar{y}_1 - b \ln x_j + \varepsilon_j, \ j = 1, 2, \ldots, N$,

ad II) $\ln Y_j = \ln A - b \ln x_j + \varepsilon_j, \ j = 1, 2, \ldots, N$,

ad IV) $\ln Y_j = \ln A - b \ln x_j + c x_j + \varepsilon_j, \ j = 1, 2, \ldots, N$.

Here, $\bar{y}_1$ is the average of the observed constituent lengths y_j of the shortest construct with the length x_1. The symbol ε_j denotes the j-th random error; Y_j means the j-th observation, its realization is y_j. We will use the general matrix form

---

[1]We followed Hřebíček's suggestion, in the private communication with the second author, to call the unit at the top of our chain of linguistic units the semantic construct, instead of the hreb or the aggregate, as documented in Andres et al. (2012a). Yet, we are open to use any other terminology in case it is proved in the further experiments that the MAL parameters are not suitable for measuring semanticity of a given sample.

for the models as well as for the estimators which are more suitable for further consideration. In particular, the matrix form of the linearized models I), II) and IV) reads $\boldsymbol{Y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where the j-th element of the vector $\boldsymbol{Y}^*$, the j-th row of the matrix $\mathbf{X}^*$ and the vector $\boldsymbol{\beta}$ of unknown regression parameters are

    ad I) $Y_j^* = \ln Y_j - \ln \bar{y}_1,\ x_j^* = -\ln x_j,\ \boldsymbol{\beta} = b,$

    ad II) $Y_j^* = \ln Y_j,\ \boldsymbol{x}_j^{*T} = (1, -\ln x_j),\ \boldsymbol{\beta} = (\ln A, b)^T,$

    ad IV) $Y_j^* = \ln Y_j,\ \boldsymbol{x}_j^{*T} = (1, -\ln x_j, x_j),\ \boldsymbol{\beta} = (\ln A, b, c)^T.$

The vector parameter $\boldsymbol{\beta}$ can be estimated either by the *ordinary least squares estimator* (OLSE), or by the *weighted least squares estimator* (WLSE) (cf. e.g. Chatterjee, Hadi, 2006, Montgomery, Peck, Vinig, 2006), in accordance with the assumptions imposed on random errors ε_j. More precisely, if the dispersion of random errors is constant, then we speak about *homoscedasticity*, otherwise, we speak about *heteroscedasticity*. For the homoscedasticity with constant dispersion σ^2, the OLSE of the vector parameter $\boldsymbol{\beta}$ is (cf. e.g. Montgomery, Peck, Vinig, 2006)

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\boldsymbol{Y}^*, \tag{6.1}$$

with the covariance matrix

$$\mathrm{var}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^{*T}\mathbf{X}^*)^{-1}, \tag{6.2}$$

where σ^2 is the unknown parameter to be estimated. The unbiased estimator of σ^2 is (cf. e.g. Efron, Tibshirani, 1993)

$$\widehat{\sigma}^2 = \frac{(\boldsymbol{Y}^* - \mathbf{X}^*\widehat{\boldsymbol{\beta}})^T(\boldsymbol{Y}^* - \mathbf{X}^*\widehat{\boldsymbol{\beta}})}{N - K}, \tag{6.3}$$

where K is the number of regression parameters (the length of the vector $\boldsymbol{\beta}$). Once the parameter σ^2 is estimated, its value can be plugged into the formula (6.2), by which the covariance matrix can be estimated as well.

For the heteroscedasticity, when the dispersion of random error ε_j is σ_j^2, $j = 1, 2, \ldots, N$, the formula for the WLSE of the vector parameter $\boldsymbol{\beta}$ takes the form

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{*T}\mathbf{W}^{-1}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{W}^{-1}\boldsymbol{Y}^*, \quad \mathrm{var}(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}^{*T}\mathbf{W}^{-1}\mathbf{X}^*)^{-1}, \tag{6.4}$$

where $\mathbf{W} = \mathrm{diag}\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2\}$ is a diagonal matrix.

If random errors are heteroscedastic, the relationship between errors and explanatory variables should be analysed in order to estimate the dispersions σ_1^2, σ_2^2, ..., σ_N^2. The algorithm is the following. Firstly, the OLSE of $\boldsymbol{\beta}$ and corresponding residual vector $\boldsymbol{e} = \boldsymbol{y}^* - \mathbf{X}^*\widehat{\boldsymbol{\beta}}$ are calculated. Next, the relationship between σ_i^2 and explanatory variables is fitted applying the ordinary least squares method to residual vector $\boldsymbol{e}$. Consequently, $\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2$ are estimated. Finally, the estimates of σ_j^2 are plugged into the matrix $\mathbf{W}$, and the WLSE of $\boldsymbol{\beta}$ can be determined. Specific situations will be discussed in detail in the next section.

If averaged models are analysed, the procedure is similar. More concretely, the matrix form of models takes the form $\overline{\boldsymbol{Y}}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where the k-th element of the

vector $\overline{\boldsymbol{Y}}^*$, the k-th row of the matrix $\mathbf{X}^*$ and the vector $\boldsymbol{\beta}$ of unknown regression parameters are

ad $\overline{\mathrm{I}}$) $\overline{Y}_k^* = \ln \overline{Y}_k - \ln \bar{y}_1$, $x_k^* = -\ln x_k$, $\beta = b$, $k = 1, 2, \ldots, n$,

ad $\overline{\mathrm{II}}$) $\overline{Y}_k^* = \ln \overline{Y}_k$, $\boldsymbol{x}_k^{*T} = (1, -\ln x_k)$, $\boldsymbol{\beta} = (\ln A, b)^T$, $k = 1, 2, \ldots, n$,

ad $\overline{\mathrm{IV}}$) $\overline{Y}_k^* = \ln \overline{Y}_k$, $\boldsymbol{x}_k^{*T} = (1, -\ln x_k, x_k)$, $\boldsymbol{\beta} = (\ln A, b, c)^T$, $k = 1, 2, \ldots, n$.

Thus, the formula for the ordinary least squares estimator of $\boldsymbol{\beta}$ is given by the equation (6.1), when the vector $\boldsymbol{Y}^*$ is replaced by $\overline{\boldsymbol{Y}}^*$. Analogously, the parameter σ^2 is unbiasedly estimated by the formula (6.3), using n instead of N. In this case, the explicit formulas for the OLSE of the parameters A, b, c are derived in Chapter 3.

When the weights $w_i = z_i / \sum_{k=1}^n z_k$, where z_i is the frequency of the i-th construct, are also taken into account in the averaged models, the vector parameter $\boldsymbol{\beta}$ should be estimated by the weighted least squares method using the expressions (6.4), where $\mathbf{W} = \mathrm{diag}\{1/w_1, 1/w_2, \ldots, 1/w_n\}$ is a diagonal matrix of the reciprocal values of weights. Let us note that the dispersions of random errors ε_k equal σ^2/w_k, and so they are not constant in this case. Thus, the covariance matrix of $\widehat{\boldsymbol{\beta}}$ is estimated by $\widehat{\sigma}^2 (\mathbf{X}^{*T} \mathbf{W}^{-1} \mathbf{X}^*)^{-1}$, where the unbiased estimator of the parameter σ^2 is given by

$$\widehat{\sigma}^2 = \frac{(\overline{\boldsymbol{Y}}^* - \mathbf{X}^* \widehat{\boldsymbol{\beta}})^T \mathbf{W}^{-1} (\overline{\boldsymbol{Y}}^* - \mathbf{X}^* \widehat{\boldsymbol{\beta}})}{n - K}. \tag{6.5}$$

Once the model is fitted, the assumptions of homoscedasticity (in the case for averaged models without weights and models with semi-averaging) and normality distribution of random errors should be tested. Homoscedasticity can be tested, e.g., by the White test (see e.g. White, 1980). Normality can be tested, e.g., by the Shapiro–Wilk test or by the Kolmogorov–Smirnov test (see e.g. Thode, 2002) applied to standardized residuals

$$\left(\frac{e_1}{\sqrt{\mathrm{var}(e_1)}}, \ldots, \frac{e_N}{\sqrt{\mathrm{var}(e_N)}} \right)^T, \qquad \mathrm{var}(\boldsymbol{e}) = \widehat{\sigma}^2 \left(\mathbf{W} - \mathbf{X}^* (\mathbf{X}^{*T} \mathbf{W}^{-1} \mathbf{X}^*)^{-1} \mathbf{X}^{*T} \right).$$

The formulas are valid for heteroscedastic models with semi-averaging. For homoscedastic models, the identity matrix is used, instead of matrix $\mathbf{W}$. When averaged models are analysed, the subscript runs from 1 to n.

If normality of random errors is not rejected, a confidence interval can be determined in a standard way by the *Wald statistic* (cf. e.g. Montgomery, Peck, Vinig, 2006). In particular, the $100(1 - \alpha)\%$ confidence interval for the parameter b is

$$I_{1-\alpha}(b) = \left[\widehat{b} - \widehat{\sigma} \sqrt{\mathrm{var}(\widehat{b})} \, t_{N-K}(1 - \alpha/2), \widehat{b} + \widehat{\sigma} \sqrt{\mathrm{var}(\widehat{b})} \, t_{N-K}(1 - \alpha/2) \right],$$

where $t_{N-K}(1 - \alpha/2)$ means the $(1 - \alpha/2)$-quantile of the Student t-distribution with $N - K$ degrees of freedom. If the parameter b is estimated in averaged models, the number N is replaced by n. The level of $100(1 - \alpha)\%$ confidence of the

confidence interval indicates the probability that the confidence range captures the true value of the parameter b.

When random errors do not manifest the normal distribution, a confidence interval for b can be determined by *bootstrap percentiles* (cf. e.g. Efron, Tibshirani, 1993). The bootstrap techniques are based on a bootstrap data set. The usual way of bootstrapping a regression model consists of bootstrapping pairs $(\boldsymbol{x}_j^*, y_j)$ from the original data set, so that a bootstrap data set $\boldsymbol{d}$ is of the form

$$\boldsymbol{d} = \{(\boldsymbol{x}_{j_1}^*, y_{j_1}), (\boldsymbol{x}_{j_2}^*, y_{j_2}), \ldots, (\boldsymbol{x}_{j_N}^*, y_{j_N})\},$$

where indices $j_1, j_2, \ldots, j_N$ represent a random sample of the integers from 1 to N. Obviously, the bootstrap data set consists of pairs $(\boldsymbol{x}_j^*, y_j)$, some appearing zero times, some appearing once, some appearing twice, etc. Let us note that bootstrap data sets should be taken from the original data set with semi-averaging.

The algorithm for determination of a bootstrap confidence interval is the following. Firstly, we generate B independent bootstrap data sets $\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_B$. For each bootstrap data set, we fit the model and compute the estimates $\widehat{b}(\boldsymbol{d}_j)$, $j = 1, 2, \ldots, B$. Then, we determine $100(\alpha/2)$-th and $100(1 - \alpha/2)$-th empirical percentiles of $\widehat{b}(\boldsymbol{d}_j)$-estimates. These percentiles will be denoted as $b_B^{\alpha/2}$ and $b_B^{1-\alpha/2}$, respectively. It means that $b_B^{\alpha/2}$ is the $B(\alpha/2)$-th value in the ordered list of values $\widehat{b}$. If $B(\alpha/2)$ is not an integer, we can take the largest integer $k \le (B+1)(\alpha/2)$ and define $b_B^{\alpha/2}$ by the k-th largest values of $\widehat{b}(\boldsymbol{d}_j)$. The resulting approximate $100(1 - \alpha)\%$ bootstrap confidence interval for the parameter b reads

$$I_{1-\alpha}(b) = \left[b_B^{\alpha/2}, b_B^{1-\alpha/2}\right].$$

The authors of the book Efron, Tibshirani (1993) suggest a general rule of thumb about the number of bootstrap replications B, for bootstrap confidence intervals, such that B should be from the interval $[500, 1000]$. Let us note that the percentile bootstrap confidence intervals are meaniful only if the bootstrap statistic has a symmetric distribution. As a simple way for verification such a symmetry, we recommend to construct a histogram of bootstrap realizations and to to check the symmetry graphically.

The goodness of fit measures (cf. e.g. Montgomery, Peck, Vinig, 2006) of a regression model widely used in practice can be characterized by means of the *residual standard error* $\widehat{\sigma}$, the *root mean square error* (*RMSE*), the *normalized root mean square error* (*NRMSE*) and the *coefficient of determination* R^2. The first three measures characterize the achieved precision fit, the latter one represents the proportion of variation explained by the model.

The quantities of *RMSE* and $\widehat{\sigma}$ can be interpreted as the average deviation of fitted and observed values of y. They can range from zero to infinity; obviously, the lower the values, the better. Since the *RMSE* is scale-dependent (*RMSE* has the same unit as the dependent variable), the application of the normalized root mean square error is more suitable. The value is often expressed as a percentage, where lower values indicate less residual variance.

The coefficient of determination R^2 tells us how accurately our model explains a phenomena. It takes values from zero to one; obviously, the higher the values, the better. In soft sciences, we believe that the threshold for a good model can start from 0.5.

The related formulas for residual standard errors are given by the square root expressions in (6.3) and (6.5). For heteroscedastic models with semi-averaging, symbol n is replaced by N in formula (6.5). Formulas for $RMSE$, $NRMSE$ and R^2 for heteroscedastic models with semi-averaging are as follows:

$$RMSE = \sqrt{\frac{(\boldsymbol{Y}^* - \mathbf{X}^*\widehat{\boldsymbol{\beta}})^T \mathbf{W}^{-1} (\boldsymbol{Y}^* - \mathbf{X}^*\widehat{\boldsymbol{\beta}})}{N}}, \qquad NRMSE = \frac{RMSE}{\boldsymbol{Y}^*_{\max} - \boldsymbol{Y}^*_{\min}},$$

$$R^2 = 1 - \frac{(\boldsymbol{Y}^* - \mathbf{X}^*\widehat{\boldsymbol{\beta}})^T \mathbf{W}^{-1} (\boldsymbol{Y}^* - \mathbf{X}^*\widehat{\boldsymbol{\beta}})}{(\boldsymbol{Y}^* - \overline{\boldsymbol{Y}}^*_W \mathbf{1})^T \mathbf{W}^{-1} (\boldsymbol{Y}^* - \overline{\boldsymbol{Y}}^*_W \mathbf{1})},$$

where $\overline{\boldsymbol{Y}}^*_W = (\mathbf{1}'\mathbf{W}^{-1}\mathbf{1})^{-1}\mathbf{1}'\mathbf{W}^{-1}$. Here, the symbol $\mathbf{1}$ denotes the vector of N units. For homoscedastic models, the identity matrix is used, instead of matrix $\mathbf{W}$. For averaged models, the symbol N is replaced by n. Note that, for cases I, $\overline{\mathrm{I}}^w$, $\overline{\mathrm{I}}$ (i.e. for the models without intercept), the value of R^2 is computed as follows:

$$R^2 = 1 - \frac{(\boldsymbol{Y}^* - \mathbf{X}^*\widehat{\boldsymbol{\beta}})^T \mathbf{W}^{-1} (\boldsymbol{Y}^* - \mathbf{X}^*\widehat{\boldsymbol{\beta}})}{\boldsymbol{Y}^{*T} \mathbf{W}^{-1} \boldsymbol{Y}^*},$$

(see e.g. Eisenhauer, 2003).

6.3 Concluding remark to a suitability of various approaches

In the case of semi-averaging, the statistical criteria have not been usually suitably satisfied. For instance, extremely low values of the coefficient of determination R^2 were obtained in our experiment from a spoken language, described in Andres, Benešová, Chvosteková, Fišerová (2014). We have also checked that the same is (even worse) true in the case with any averaging. Moreover, in the latter case, the integer-valued constituent lengths would require a completely different statistical methodology. This suggests to apply preferably the averaged formulas.

Peter Angermann, *A. I. (Nohandsfractal)*, 2001, oil on canvas

7 EXPERIMENT 1: TOWARDS A FRACTAL ANALYSIS OF THE SIGN LANGUAGE

7.1 Formulation of goal

Hřebíček formulated two versions of the conjecture about a fractal structure of languages. The weak version simply says that *the Menzerath–Altmann law (MAL), formulated by Altmann* (see Altmann, 1980), holds for all language levels, including supra-sentence levels like semantic constructs (for their definition, see Hřebíček, 1992). In other words, language structures exhibit *a self-similarity property* with respect to the validity of MAL on each level, up to the parameters which characterize a given language structure. The strong version is based on *an isomorphism between the logarithmized form of MAL and the Moran–Hutchinson formula* (see e.g. Barnsley, 2013) for the calculation of the fractal dimension of strictly self-similar objects.

Relaxing the regularity of the lexicographic trees and the strict self-similarity property to the block regularity and the block self-similarity, where the self-similarity property holds in blocks, but not necessarily on each level, we were able to generalize the formula for the calculation of the fractal dimension of *language fractals* (i.e. texts for which all the "shape" parameters of MAL are positive). The calculated dimension is called there a *measure of semanticity* of a given text (see Chapter 4). For some further information about the fractality of spoken language texts, see e.g. Andres, Benešová, Kubáček, Vrbková (2012a), Ausloos (2010, 2012), Drożdż et al. (2016), Köhler (1995, 1997, 2008, 2014) and Montemurro, Pury (2002).

It is a natural challenge to ask whether or not a fractal structure conjecture can be extended from spoken languages to a sign language? There is an enormous amount of the literature about the analysis of a sign language, including the quantitative analysis (see e.g. Frosini, 1991; Frosini, Landi, 2001; Handouyahia, Ziou, Wang, 1999; Johnston, Schembri, 2007; Joshua, Borneman, Malaia, Wilbur, 2018; Malaia, 2017; Malaia, Borneman, Wilbur, 2016, 2018; Stewart, 2014; Uras, Verri, 1994, 1995, and the references therein). In particular, the fractal complexity investigation in the recent papers by Joshua, Borneman, Malaia, Wilbur (2018) and Malaia, Borneman, Wilbur (2016, 2018) is along the lines of our present

research aim. On the other hand, despite a further isomorphism between the applied mathematical formulas, our tools and approaches are completely different.

This chapter is organized as follows. At first, it will be convenient to collect some basic preliminaries from a sign language in comparison to spoken languages. Then the applied mathematical and statistical elements will be briefly recalled. The obtained results of our analysis will be presented in the form of tables and graphs. Finally, they will be discussed and commented in concluding remarks.

7.2   Basic preliminaries from sign language

Sign languages differ from vocal spoken languages, besides other things, by the mode of their existence. In particular, unlike audio-oral spoken languages, sign languages are of a visual-motor nature. Moreover, unlike de Saussure's principle of linearity holding in principle for spoken languages (see e.g. Andres, 2009, and the references therein), the character of sign languages is obviously simultaneous (multidimensional). Nevertheless, it is usually expected that sign languages of the deaf have all the necessary characteristics of natural languages and are therefore full-fledged (cf. Stokoe, 1960). The nature of a sign language does not however reflect the character of natural languages in many aspects like the mentioned multidimensionality vs one-dimensionality (i.e. linearity) of spoken languages as well as their hierarchical structures.

Schematically, the comparison of the hierarchical structure of spoken and sign languages can be roughly expressed in Table 7.1.

As concerns our experiments, the research data of the Czech sign language were collected by the usual elicitation method, i.e. solely from native speakers of Czech sign language in the shape of monologues on given topics like: my hobbies, last holiday, my family, my job, etc., in natural environments. Three thoroughly selected native deaf speakers of Czech sign language supervised the grammar and lexical levels of 31 recorded speeches which were approximately 10 minutes long. They have been recorded in Czech sign language from various respondents and filmed by means of digital video cameras with high resolution.

Table 7.1: Language levels in the text/speech (constructs and constituents)

Spoken language	Sign language
semantic	construct ?
sentence	sentence
clause	clause
word	sign
syllable	?
phoneme (letter)	? (phoneme/morpheme)

For the annotation and segmentation of research data, the application ELAN language annotator was used. In this application, it is also possible to define the lengths of individual constituents, when using the timeline. Segmentation was carried out by deaf native Czech sign language speakers under the supervision of a deaf supervisor with linguistic education. The detection of language units (sentences, clauses, signs) was based on the empirical experience of native speakers of Czech sign language in analogy with spoken languages (see Altmann, 1980; Andres, Benešová, Kubáček, Vrbková, 2012; Hřebíček, 1992).

Along the lines of the ideas of Köhler (2014), there are two ways of counting the lengths of constituents. The first one is to count the number of individual units of a given level. In the second, one calculates the time lengths. For a given text, these two techniques (Köhler speaks with this respect about "motifs") can also be confirmed on different levels.

7.3   Applied formulas of MAL

As a main tool, we will employ the Menzerath–Altmann law (MAL).

Heuristically, it says that *the longer a language construct is, the shorter its constituents are*. Geometrically, the related *graphs* of the respective functions, expressing the dependence between constructs and constituents, *should be decreasing*. Mathematically, the *complete formula* of MAL takes the form (see Chapter 3)

$$y = A\,x^{-b}\mathrm{e}^{cx}, \tag{7.1}$$

or, equivalently (for $b \neq 0$),

$$\frac{1}{b} = \frac{\log x}{\log\left(\frac{A}{y}\exp(cx)\right)},$$

where A, b, c are real parameters, $\mathrm{e} = 2.718\ldots$ is the Euler number, x are natural numbers and y are real numbers.

For $c = 0$, it reduces to the *truncated formula*, namely

$$y = Ax^{-b}, \quad \text{resp. (for } b \neq 0\text{)}, \ \frac{1}{b} = \frac{\log x}{\log\left(\frac{A}{y}\right)}. \tag{7.2}$$

In the particular case, when $A\mathrm{e}^c = y(1) = y_1$, we get the formula

$$y = y_1 x^{-b}\mathrm{e}^{c(x-1)}, \quad \text{resp. (for } b \neq 0\text{)}, \ \frac{1}{b} = \frac{\log x}{\log\left(\frac{y_1}{y}\exp(cx - \mathrm{e})\right)}, \tag{7.3}$$

which for $c = 0$ reduces to

$$y = y_1 x^{-b}, \quad \text{resp. (for } b \neq 0\text{)}, \ \frac{1}{b} = \frac{\log x}{\log\frac{y_1}{y}}. \tag{7.4}$$

The most important parameter b, which should be according to heuristic as well as geometric versions positive, is usually called a *shape parameter*. For its numerical calculations, statistical verifications and optimal usage of the above formulas, see Chapters 3 and 6.

Our task will be so to detect the situations in which the calculated shape coefficient b is positive. We will also have to verify the obtained results statistically, by means of the following standard criteria (see Chapter 6):

- The *root mean square error* (RMSE) and the *normalized root mean square error* (NRMSE).
- The *coefficient of determination* $\mathbb{R}^2 \in [0, 1]$.
- *Homoscedasticity*.
- *Normality*.

If we are able to consider n (sign) language levels and if all the shape parameters b_i, $i = 1, \ldots, n$ (the index i indicates the appropriate level) are there positive, i.e. $b_i > 0$, for every $i = 1, \ldots, n$, then we can speak about a *(sign) language fractal*. Its *fractal dimension*

$$D = \frac{n}{b_1 + \cdots + b_n}, \tag{7.5}$$

i.e. the reciprocal value of the arithmetic average of all shape parameters, is called the *measure of semanticity* of a (sign) language fractal. For more details, see Chapter 4.

7.4   Results of analysis

We will consider two sign language levels (sentences vs clauses and clauses vs signs). The lengths of level units will be measured in three, resp. two different ways (cases 1, 2, 3). All the formulas (7.1)–(7.4) of MAL from Section 7.3 will be applied, where formula (7.1) corresponds to model 4, formula (7.2) to model 2, formula (7.3) to model 3 and formula (7.4) to model 1. The values of constituents will be averaged in an arithmetic way. The relative frequencies of constructs, called weights, will be implemented into calculations in an extra way (i.e. alternatively to calculations without their implementation), which is briefly indicated as "weighted".

Level 1 – length of sentences in constituents vs average length of clauses in subconstituents

The three cases in Table 7.2 will be under our consideration.

Table 7.2: Overview of units of constructs and constituents at Level 1

Case	Construct (in constituents)	Constituent – averaged (in subconstituents)
1	sentence (in number of clauses)	clause (in number of signs)
2	sentence (in seconds)	clause (in seconds)
3	sentence (in number of clauses)	clause (in seconds)

Level 1, Case 1

Table 7.3: The relationship between the lengths of sentences *x* (measured in the number of clauses), occurring with the frequency of *z*, and the average lengths of clauses *y* (measured in the number of signs)

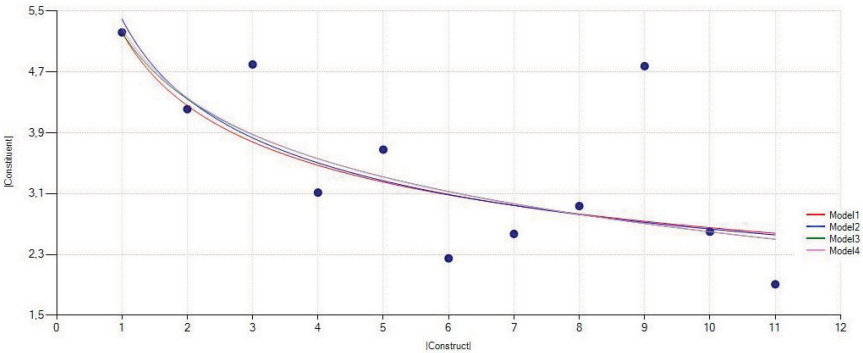| x	y	z	x	y	z	x	y	z
1	5.2222	9	5	3.6800	50	9	4.7778	9
2	4.2121	66	6	2.2500	12	10	2.6000	10
3	4.8000	15	7	2.5714	7	11	1.9091	11
4	3.1154	52	8	2.9375	16			

Figure 7.1: The graph illustrating the MAL curve calculated from the input data presented in Table 7.3
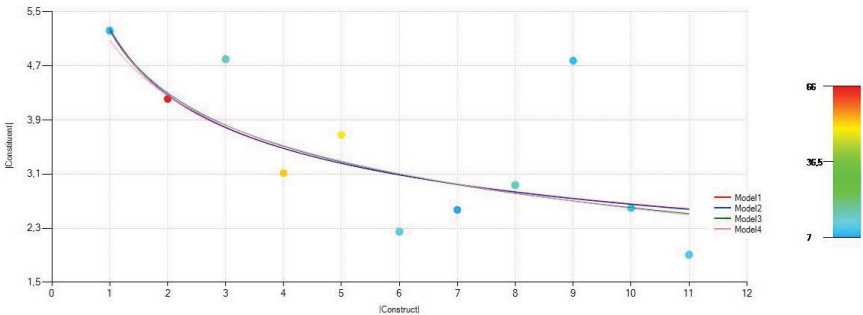
Figure 7.2 (weighted): The graph illustrating the MAL curve calculated from the input data presented in Table 7.3

Table 7.4: Statistical evaluation of the experiment at Level 1, Case 1; A, b, c are the parameters of MAL, (normalized) root mean square error, R^2 is the coefficient of determination, homoscedasticity and normality

	A	b	c	RMSE	NRMSE	R^2	Homosce-dasticity	Normality
Model 1		0.2940		0.2345	0.2330	0.8268	not rejected	not rejected
Model 2	5.3995	0.3115		0.2341	0.2326	0.4719	not rejected	not rejected
Model 3		0.2426	-0.0155	0.2337	0.2323	0.8279	not rejected	not rejected
Model 4	5.3472	0.2544	-0.0135	0.2337	0.2323	0.4736	not rejected	not rejected

Table 7.5: Statistical evaluation of the experiment at Level 1, Case 1 (weighted); A, b, c are the parameters of MAL, (normalized) root mean square error, R^2 is the coefficient of determination, homoscedasticity and normality

	A	b	c	RMSE	NRMSE	R^2	Homosce-dasticity	Normality
Model 1		0.2929		0.8189	0.8138	0.8695	not rejected	not rejected
Model 2	5.2580	0.2971		0.8188	0.8137	0.5188	not rejected	not rejected
Model 3		0.2664	-0.0091	0.8173	0.8122	0.8701	not rejected	not rejected
Model 4	5.1629	0.2307	-0.0157	0.8165	0.8114	0.5216	not rejected	not rejected

Level 1, Case 2

Figure 7.3 (weighted): The graph illustrating the MAL curve calculated from the input data presented in Table 7.6

Table 7.6: The relationship between the lengths of sentences x (measured in seconds), occurring with the frequency of z, and the average lengths of clauses y (measured in seconds)

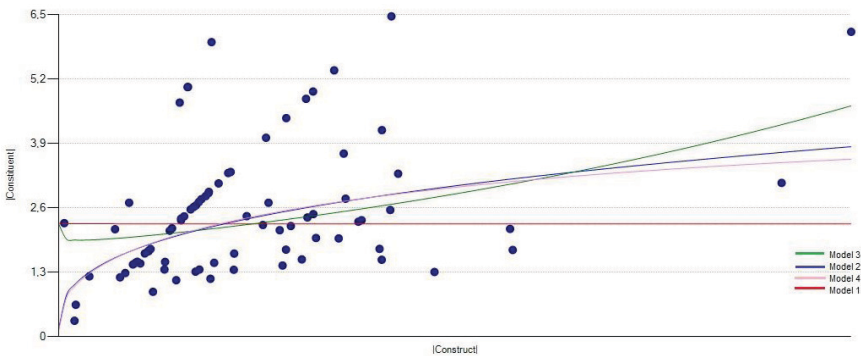| x	y	z	x	y	z	x	y	z	x	y	z
0.6300	2	0.3180	4.4310	2	2.1875	6.0730	4	1.4890	10.0410	5	1.9908
0.6800	1	0.6410	4.5940	4	1.1380	6.2450	2	3.0950	10.7490	2	5.3810
1.2080	1	1.2150	4.7300	1	4.7290	6.6190	2	3.3050	10.9260	5	1.9822
2.2090	1	2.1690	4.7820	2	2.3500	6.7160	2	3.3260	11.1220	3	3.6977
2.3200	1	2.2920	4.7840	2	2.3830	6.8380	5	1.3484	11.1950	4	2.7850
2.4050	2	1.1960	4.9040	2	2.4290	6.8530	4	1.6760	11.6930	5	2.3194
2.6070	2	1.2820	5.0350	1	5.0420	7.3420	3	2.4337	11.8200	5	2.3514
2.7600	1	2.7040	5.0550	1	5.0440	7.9730	7	2.2543	12.5190	7	1.7711
2.9080	2	1.4560	5.1560	2	2.5690	8.0940	2	4.0175	12.6030	8	1.5509
3.0040	2	1.4875	5.2750	2	2.6185	8.1910	3	2.7040	12.610	3	4.1720
3.0750	2	1.5115	5.3460	4	1.3103	8.6280	4	2.1483	12.9350	5	2.5572
3.1950	2	1.4765	5.3650	2	2.6510	8.7380	6	1.4350	12.9790	2	6.4725
3.3690	2	1.6790	5.4720	2	2.7170	8.8720	5	1.7552	13.2430	4	3.2905
3.5000	2	1.7230	5.4960	4	1.3533	8.8830	2	4.4145	14.6580	11	1.3043
3.5220	2	1.7265	5.5740	2	2.7755	9.0600	4	2.2333	17.6040	8	2.1756
3.5940	2	1.7680	5.7430	2	2.8380	9.4880	6	1.5590	17.7070	10	1.7490
3.6870	4	0.9033	5.8510	2	2.8995	9.6530	2	4.8050	28.1860	9	3.1050
4.1360	3	1.3543	5.8660	2	2.9230	9.7070	4	2.4065	30.8980	5	6.1586
4.1600	2	1.5085	5.9290	5	1.1672	9.9270	2	4.9525			
4.3400	2	2.1380	5.9700	1	5.9510	9.9370	4	2.4745			
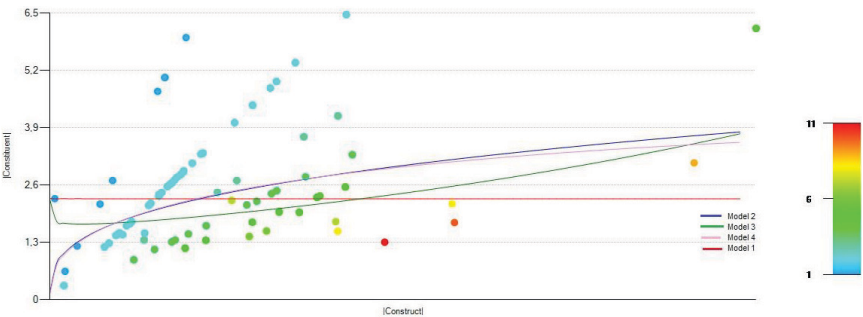
Figure 7.4 (weighted): The graph illustrating the MAL curve calculated from the input data presented in Table 7.6

Table 7.7: Statistical evaluation of the experiment at Level 1, Case 2; A, b, c are the parameters of MAL, (normalized) root mean square error, R^2 is the coefficient of determination, homoscedasticity and normality

	A	b	c	RMSE	NRMSE	R^2	Homosce-dasticity	Normality
Model 1		0.0008		2.1344	0.7083	0.0002	rejected	not rejected
Model 2	0.1217	-0.3338		1.6725	0.5551	0.2290	rejected	not rejected
Model 3		0.0282	0.0000	1.7970	0.5963	0.1355	rejected	not rejected
Model 4	0.0943	-0.3677	0.0000	1.6819	0.5582	0.2292	rejected	not rejected

Table 7.8: Statistical evaluation of the experiment at Level 1, Case 2 (weighted); A, b, c are the parameters of MAL, (normalized) root mean square error, R^2 is the coefficient of determination, homoscedasticity and normality

	A	b	c	RMSE	NRMSE	R^2	Homosce-dasticity	Normality
Model 1		0.0001		1.8814	0.6244	0.0000	rejected	not rejected
Model 2	0.1218	-0.3337		1.8641	0.6186	0.2267	rejected	not rejected
Model 3		0.0463	0.0000	0.7642	0.2536	0.1834	not rejected	not rejected
Model 4	0.0956	-0.3660	0.0000	1.8868	0.6262	0.2312	rejected	not rejected

Level 1, Case 3

Table 7.9: The relationship between the lengths of sentences x (measured in the number of clauses), occurring with the frequency of z, and the average lengths of clauses y (measured in seconds)

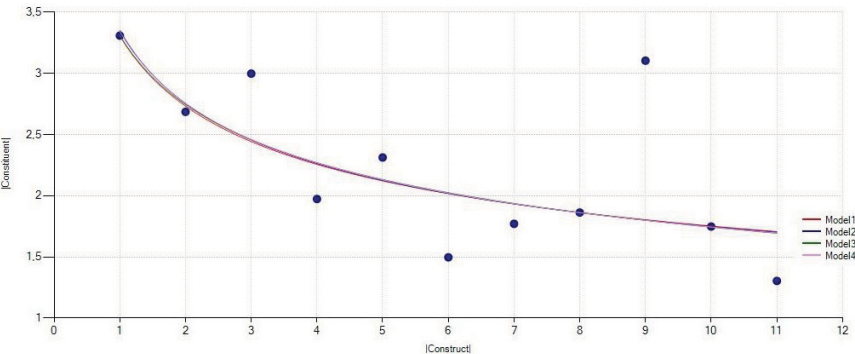| x	y	z	x	y	z	x	y	z
1	3.3097	9	5	2.3134	50	9	3.1050	9
2	2.6865	66	6	1.4970	12	10	1.7490	10
3	2.9991	15	7	1.7711	7	11	1.3043	11
4	1.9739	52	8	1.8633	16			

Figure 7.5: The graph illustrating the MAL curve calculated from the input data presented in Table 7.9

Table 7.10: Statistical evaluation of the experiment at Level 1, Case 3; A, b, c are the parameters of MAL, (normalized) root mean square error, R^2 is the coefficient of determination, homoscedasticity and normality

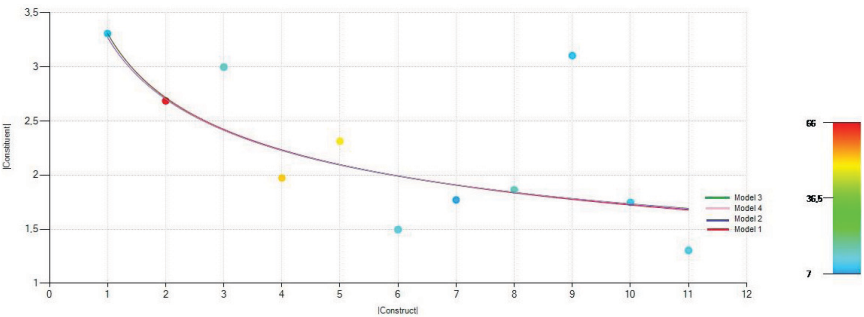|	A	b	c	RMSE	NRMSE	R^2	Homosce-dasticity	Normality
Model 1		0.2761		0.2198	0.2361	0.8272	not rejected	not rejected
Model 2	3.3526	0.2828		0.2198	0.2360	0.4552	not rejected	not rejected
Model 3		0.2637	-0.0037	0.2198	0.2360	0.8273	not rejected	not rejected
Model 4	3.3481	0.2750	-0.0019	0.2198	0.2360	0.4553	not rejected	not rejected

Figure 7.6 (weighted): The graph illustrating the MAL curve calculated from the input data presented in Table 7.9

Table 7.11: Statistical evaluation of the experiment at Level 1, Case 3 (weighted); A, b, c are the parameters of MAL, (normalized) root mean square error, R^2 is the coefficient of determination, homoscedasticity and normality

	A	b	c	RMSE	NRMSE	R^2	Homosce-dasticity	Normality
Model 1		0.2834		0.7686	0.8253	0.8763	not rejected	not rejected
Model 2	3.2726	0.2765		0.7683	0.8250	0.5148	not rejected	not rejected
Model 3		0.2927	0.0032	0.7683	0.8251	0.8764	not rejected	not rejected
Model 4	3.2762	0.2805	0.0009	0.7682	0.8250	0.5148	not rejected	not rejected

Level 2 – length of clauses in constituents vs average length of signs in sub-constituents

The same procedure has been applied to the relationship between clauses and signs.

Table 7.12: Overview of units of constructs and constituents at Level 2

Case	Construct (in constituents)	Constituent – averaged (in subconstituents)
1	clause (in number of signs)	signs (in number of phonemes/morphemes?)
2	clause (in seconds)	signs (in seconds)
3	clause (in number of signs)	signs (in seconds)

Level 2, Case 1

In view of the above-mentioned simultaneity (multidimensionality) of sign languages (see Section 2), it is difficult to identify the number of constituents composing signs as the smallest semantic units of a sign language for the purposes of quantitative analysis. The analogy of syllables (for the words of spoken languages) could not be identified successfully. So we were forced to skip this level completely and focus on the lowest level of phonemes, respectively morphemes (see Langer, Rypka, 2017). Due to the extremely long time needed for this segmentation, the resulting insufficient amount of data and because of the fact that the calculated values do not correspond to MAL, we decided to omit this analysis here.

Level 2, Case 2

Since the table for the relationship between the lengths of clauses x (measured in seconds), occurring with the frequency of z, and the average lengths of signs y (measured in seconds) is rather huge (253 lines) and especially since the obtained results are negative for all four models (see Tables 7.13 and 7.14 below), we decided not to present the related data here.
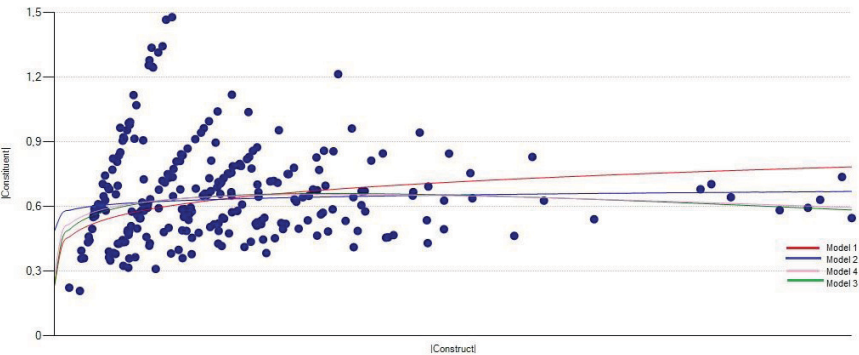

Figure 7.7: The graph illustrating the MAL curve calculated from the non-presented data related to Level 2, Case 2

Table 7.13: Statistical evaluation of the experiment at Level 2, Case 2; A, b, c are the parameters of MAL, (normalized) root mean square error, R^2 is the coefficient of determination, homoscedasticity and normality

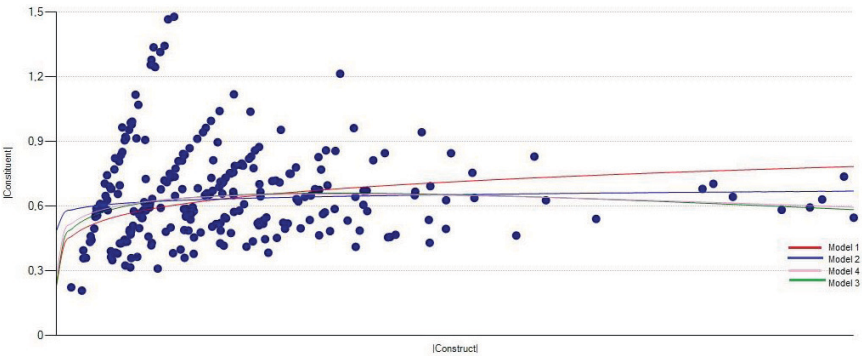|	A	b	c	RMSE	NRMSE	R^2	Homosce-dasticity	Normality
Model 1		-0.1361		1.8974	0.9670	0.9183	rejected	not rejected
Model 2	0.4825	-0.0355		1.8926	0.9645	0.0072	rejected	not rejected
Model 3		-0.1511	0.0000	1.9429	0.9901	0.9306	rejected	not rejected
Model 4	0.2753	-0.1207	0.0000	1.9627	1.0002	0.0149	rejected	not rejected

Figure 7.8 (weighted): The graph illustrating the MAL curve calculated from the non-presented data related to Level 2, Case 2

Table 7.14: Statistical evaluation of the experiment at Level 2, Case 2 (weighted); A, b, c are the parameters of MAL, (normalized) root mean square error, R^2 is the coefficient of determination, homoscedasticity and normality

	A	b	c	RMSE	NRMSE	R^2	Homosce-dasticity	Normality
Model 1		-0.1361		1.8790	0.9576	0.9184	rejected	not rejected
Model 2	0.4797	-0.0363		1.8905	0.9634	0.0075	rejected	not rejected
Model 3		-0.1511	0.0000	1.9433	0.9903	0.9308	rejected	not rejected
Model 4	0.2728	-0.1220	0.0000	1.9477	0.9926	0.0153	rejected	not rejected

Level 2, Case 3

Table 7.15: The relationship between the lengths of clauses x (measured in the number of signs), occurring with the frequency of z, and the average lengths of signs y (measured in seconds)

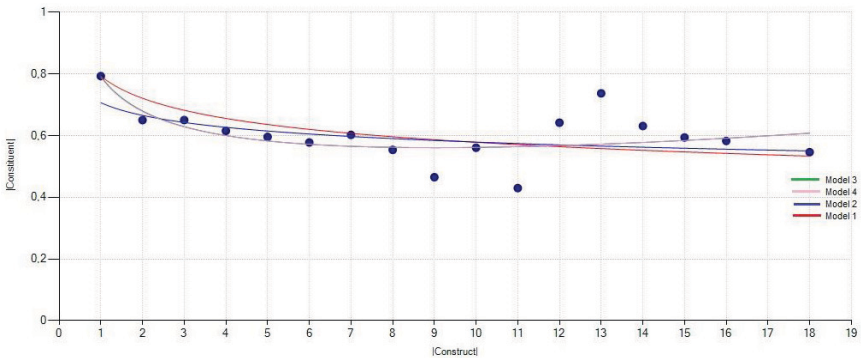| x	y	z	x	y	z	x	y	z
1	56	0.7934	7	8	0.6026	13	1	0.7375
2	57	0.6506	8	4	0.5540	14	1	0.6314
3	48	0.6509	9	5	0.4655	15	1	0.5943
4	29	0.6157	10	2	0.5611	16	1	0.5831
5	23	0.5962	11	1	0.4301	18	1	0.5467
6	15	0.5781	12	4	0.6423			

Figure 7.9: The graph illustrating the MAL curve calculated from the input data presented in Table 7.15

Table 7.16: Statistical evaluation of the experiment at Level 2, Case 3; A, b, c are the parameters of MAL, (normalized) root mean square error, R^2 is the coefficient of determination, homoscedasticity and normality

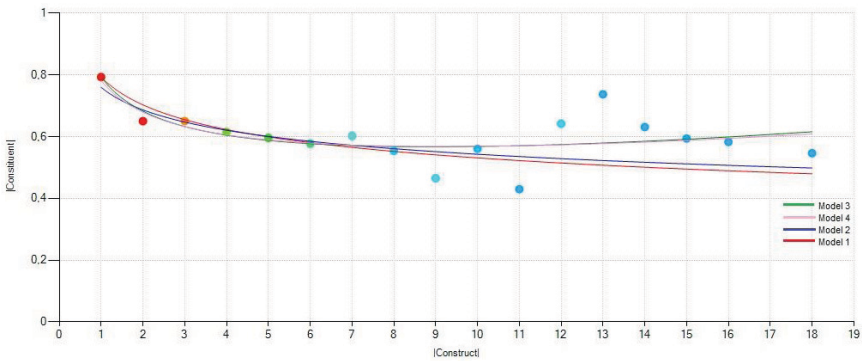|	A	b	c	RMSE	NRMSE	R^2	Homosce-dasticity	Normality
Model 1		0.1370		0.1310	0.2139	0.8312	not rejected	not rejected
Model 2	0.7071	0.0865		0.1240	0.2026	0.2265	not rejected	not rejected
Model 3		0.2641	0.0293	0.1128	0.1842	0.8748	not rejected	not rejected
Model 4	0.7667	0.2594	0.0288	0.1128	0.1842	0.3606	not rejected	not rejected

Figure 7.10 (weighted): The graph illustrating the MAL curve calculated from the input data presented in Table 7.15

Table 7.17: Statistical evaluation of the experiment at Level 2, Case 3 (weighted); A, b, c are the parameters of MAL, (normalized) root mean square error, R^2 is the coefficient of determination, homoscedasticity and normality

	A	b	c	RMSE	NRMSE	R^2	Homosce-dasticity	Normality
Model 1		0.1740		0.2440	0.3985	0.9226	not rejected	not rejected
Model 2	0.7600	0.1458		0.2242	0.3662	0.7664	not rejected	not rejected
Model 3		0.2582	0.0290	0.1901	0.3104	0.9530	not rejected	not rejected
Model 4	0.7619	0.2406	0.0262	0.1880	0.3070	0.8358	not rejected	not rejected

7.5   Discussion on experiments

Summing up the calculations presented in tables above, we could see that the positive results (with positive values of a shape parameter b) were obtained in the following cases:

- Level 1, Case 1 (see Table 7.4),
- Level 1, Case 1 – weighted (see Table 7.5),
- Level 1, Case 3 (see Table 7.8),
- Level 1, Case 3 – weighted (see Table 7.11),
- Level 2, Case 3 (see Table 7.16),
- Level 2, Case 3 – weighted (see Table 7.17).

Otherwise, the obtained results were negative (see Tables 7.7, 7.8, 7.13 and 7.14). Let us note that, despite the positive values of b, calculated for models 1 and 3 at Level 1, Case 3 (see Tables 7.7 and 7.8), their statistical verification failed. Moreover, the related graphs in Figures 7.1 and 7.4 were not convincing.

Unfortunately, we were only able to consider two sign language levels. The detection of analogies of supra-sentence levels or lower levels like those of words

vs syllables, indicated in Table 7.1, seems to be rather ineffective and cumbersome. We tried to investigate for the needs of fractal analysis the third level of signs vs "something" what can be characterized as the compositions of phonemes (i.e. supra-phonemes), but so far with no success. Nevertheless, the *sign language fractals*, defined for $n = 2$ levels in Section 7.3, jointly with the associated *measure of semanticity*

$$D = \frac{2}{b_1 + b_2} \qquad \text{(see (7.5))},$$

can be collected in Table 7.18. Observe that $D \in [3.6304, 5.4157]$ which is a bit less, but not far from our experience with spoken language texts.

Table 7.18: The measure of semanticity D of sign language fractals; the abbreviation w. denotes "weighted"

	Level 1, Case 1 Level 2, Case 3	Level 1, Case 3 Level 2, Case 3	Level 1, Case 1 w. Level 2, Case 3 w.	Level 1, Case 3 w. Level 2, Case 3 w.
Model 1	4.6404	4.8414	4.2836	4.3725
Model 2	5.0251	5.4157	4.5157	4.7360
Model 3	3.9471	3.7893	3.8124	3.6304
Model 4	3.8926	3.7425	4.2436	3.8380

After all, since our experiments were the first attempt in this research field, the obtained results must be considered only in a preliminary way. Nevertheless, these preliminary indicators signalize that because of closer analogies at higher levels (described in Table 7.1), the verification of MAL could be expected just there. On the other hand, perhaps mainly due to the multidimensional character of signs (in contrast to a linearity of spoken languages), their one-dimensional "projections" into the length, resp. time, dimensions might have been too restrictive. The comparative fractal analysis of spoken and sign languages texts on the same topic will be treated by ourselves elsewhere.

7.6   Concluding remarks

In order to understand the fractal approach in a deeper way, let us recall the *Moran–Hutchinson formula*, namely (see e.g. Barnsley, 2013)

$$D = \frac{\log m}{\log \frac{1}{r}}, \tag{7.6}$$

for the calculation of the (self-similarity) fractal dimension of a given fractal structure, where m denotes the number of its parts on each scale (level) and $r \in (0, 1)$ is a related factor of contraction, provided the given structure is either totally disconnected or its parts are at most touching.

Thus, there is an evident isomorphism between (7.6) and every logarithmic form of MAL. For instance, for the truncated formula (7.2), this isomorphism takes the form of correspondence ($A > 0$, $b > 0$)

$$D \sim \frac{1}{b}, \quad m \sim x, \quad r \sim \frac{y}{A} \ (= x^{-b}).$$

The first correspondence $D \sim 1/b$ is evidently a special case of (7.5) for $b = b_1 = \cdots = b_n$.

Let us finally note that formally the same formula as (7.2) and (7.6) occurs also in the recent papers of Joshua, Borneman, Malaia, Wilbur (2018) and Malaia, Borneman, Wilbur (2016, 2018). More precisely, their formula takes the form

$$M(f) = \alpha f^{-\beta}$$

where M is a magnitude of optical flow, f stands for the frequency, α is a fitting parameter for the spectral density amplitude and β is a fitting parameter for the "fractal complexity". The above correspondences can therefore be extended as follows:

$$D \sim \frac{1}{b} \sim \frac{1}{\beta}, \quad m \sim x \sim f, \quad r \sim \frac{y}{A} \ (= x^{-b}) \sim \frac{M}{\alpha} \ (= f^{-\beta}).$$

One can readily check that the "fractal complexity", expressed in terms of a (self-similarity) fractal dimension, is then rather curiously a reciprocal value of D. On this basis, we get (for $\beta > 0$) immediately that

$$\frac{1}{\beta} = \frac{\log f}{\log \frac{\alpha}{M}}. \tag{7.7}$$

It would be certainly interesting to compare the calculated values of D, obtained via our approach by means of MAL, and the one by means of (7.7) (i.e. $D \sim 1/\beta$), in order to detect a possible mutual relationship.

Peter Angermann, *A. I. (Nohandsfractal)*, 2010, acrylic on canvas

8 EXPERIMENT 2: APPLICATION OF THE HURST FORMULA

8.1 Formulation of problem

The main theoretical aim of the present chapter is to use in a "block-like" way an isomorphism between the well known Moran–Hutchinson formula for the dimension of self-similar fractals (see Chapter 4) and the Hurst formula for a measure of a long-term memory of time series (see Hurst, 1951; Hurst, Black, Simaika, 1965; Mandelbrot, Wallis, 1968; Mosteller, Wallace, 1964), in order to analyse quantitatively sign language structures. By a "block-like" way, we again mean that the self-similarity is not necessarily assumed on each scale but, more naturally, for the whole blocks of several (in our experiment, three) sign language levels.

As already pointed out, a quantitative approach to sign languages is, unlike to spoken languages, still rather rare (see e.g. Andres, Benešová, Langer, 2019; Joshua et al., 2018; Malaia, Borneman, Wilbur, 2018, and the references therein), because it requires delicate techniques.

This chapter is organized as follows. At first, we recall Hutchinson's idea (see e.g. Barnsley, 2013; Hutchinson, 1981) to construct fractals as attractors of the iterated function systems (IFS), the Moran–Hutchinson formula for dimension of self-similar fractals (see e.g. Hutchinson, 1981; Moran, 1946), and their modification for cyclically (block-like) self-similar fractals (see e.g. Andres, 2014; Andres, Rypka, 2012). Then the relationship between the Hurst formula (see e.g. Hurst, 1951; Hurst, Black, Simaika, 1965; Mosteller, Wallace, 1964) and the Moran–Hutchinson formula is examined for a cyclically repeated family of time series. According to the resulting main theorem, the respective Hausdorff dimension equals approximately the reciprocal value of an arithmetic mean of the numerically calculated and statistically verified Hurst exponents on scaling levels. Before applying this theorem (as an illustrative example) to a given sign language text, we emphasize some specifica of the related sign language aspects of quantitative linguistics. The obtained results are interpreted and discussed just in these terms.

8.2   Dimension of cyclically (block-like) self-similar fractals

The following seminal idea to construct fractals as attractors of the iterated function systems (IFS) was originated by Hutchinson (1981) and popularized by Barnsley (2013).

Let (X,d) be a complete metric space and $f_i\colon X \to X$, $i = 1,\ldots,n$, be the contractions with factors $r_i \in [0,1)$, $i = 1,\ldots,n$, i.e.

$$d\left(f_i(x),f_i(y)\right) \leq r_i d(x,y), \text{ for all } x,y \in X \text{ and } i = 1,\ldots,n.$$

The collection $\{(X,d), f_i\colon X \to X, i = 1,\ldots,n\}$ is then called an *iterated function system (IFS)*.

Let $(K(X),d_H)$ be a class of all compact subsets of X endowed with the Hausdorff metric d_H, i.e.

$$d_H(A,B) := \max\left\{\sup_{a \in A} d(a,B), \sup_{b \in B} d(b,A)\right\}$$

$$= \max\left\{\sup_{a \in A}\left(\inf_{b \in B} d(a,b)\right), \sup_{b \in B}\left(\inf_{a \in A} d(a,b)\right)\right\},$$

where $A,B \in K(X)$, which is called the *hyperspace* associated with (X,d).

It is well known (see e.g. Barnsley, 2013, Hutchinson, 1981) that this hyperspace $(K(X),d_H)$ is also complete and the induced *hypermaps* $f_i^*\colon K(X) \to K(X)$, where

$$f_i^*(A) := \bigcup_{x \in A} f_i(x), \quad A \in K(X), \ i = 1,\ldots,n,$$

are also contractions with factors $r_i \in [0,1)$, $i = 1,\ldots,n$, i.e.

$$d_H\left(f_i^*(A),f_i^*(B)\right) \leq r_i d_H(A,B), \text{ for all } A,B \in K(X) \text{ and } i = 1,\ldots,n.$$

Moreover, the *Hutchinson–Barnsley operator* $\mathscr{F}\colon K(X) \to K(X)$, where

$$\mathscr{F}(A) := \bigcup_{i=1}^{n} f_i^*(A), \quad A \in K(X), \tag{8.1}$$

is a contraction with the factor $r = \max_{i=1,\ldots,n} r_i$, i.e.

$$d_H(\mathscr{F}(A),\mathscr{F}(B)) \leq r d(A,B), \quad \text{for all } A,B \in K(X). \tag{8.2}$$

Hence, applying the well known Banach contraction principle, there exists a unique fixed point $A^* \in K(X)$ such that $A^* = \mathscr{F}(A^*)$ and

$$\lim_{k \to \infty} d_H\left(A^*,\mathscr{F}^k(A_0)\right) = 0, \quad \text{for every } A_0 \in K(X), \tag{8.3}$$

where $\mathscr{F}^k(A_0) = \mathscr{F} \circ \mathscr{F}^{k-1}(A_0)$, $\mathscr{F}^0(A_0) = A_0$, jointly with

$$d_H\left(A^*,\mathscr{F}^k(A_0)\right) \leq \frac{r^k}{1-r} d_H(\mathscr{F}(A_0),A_0), \quad k = 1,2,\ldots \tag{8.4}$$

The fixed point $A^* \in K(X)$ of $\mathscr{F}$, i.e. $A^* = \mathscr{F}(A^*)$, which is in view of (8.3) an *attractor* of the given IFS, we call a *fractal*.

Its *similarity dimension* $\dim A^*$ can be estimated from above by means of the *Moran–Hutchinson formula* (see Chapter 4):

$$\sum_{i=1}^n r_i^D = 1, \tag{8.5}$$

i.e. $\dim A^* \leq D$, provided the *open set condition* is satisfied, namely that there exists an open set $V \subset X$ such that $\mathscr{F}(V) = V$ and

$$f_i(V) \cap f_j(V) = \emptyset, \quad i \neq j; \; i, j = 1, \dots, n.$$

Observe that for $r = r_1 = \cdots = r_n$, formula (8.5) takes the particular form

$$D = \frac{\log n}{\log \frac{1}{r}}. \tag{8.6}$$

In case of IFS consisting of similitudes, the equality $D = \dim A^*$ holds. In Euclidean spaces $X = \mathbb{R}^k$, the similarity dimension equals the *Hausdorff dimension* which will be convenient for our application.

Summing up, we can formulate the Hutchinson theorem in the form of the following proposition.

Proposition 8.1 (cf. Barnsley, 2013; Hutchinson, 1981)**.** *For every IFS $\{(X,d),$ $f_i \colon X \to X, i = 1, \dots, n\}$ consisting of similitudes, there exists a unique compact invariant set $A^* \subset X$ of the (multivalued) operator $\bigcup_{i=1}^n f_i \colon X \to K(X)$, i.e.*

$$A^* = \bigcup_{i=1}^n \bigcup_{x \in A^*} f_i(x),$$

which satisfies relations (8.3), (8.4). Its fractal (similarity) dimension $D = \dim A^*$ can be calculated by means of formula (8.5), provided the open set condition is satisfied.

Now, consider the IFS associated with the composition

$$\mathscr{F}_m \circ \cdots \circ \mathscr{F}_1 \colon K(X) \to K(X)$$

of m Hutchinson–Barnsley operators $\mathscr{F}_j \colon K(X) \to K(X)$, $j = 1, \dots, m$, where (cf. (8.1), (8.2))

$$\mathscr{F}_j(A) := \bigcup_{i=1}^n f_{i,j}^*(A), \quad A \in K(X), \tag{8.7}$$

$$d_H\left(\mathscr{F}_j(A), \mathscr{F}_j(B)\right) \leq s_j d_H(A, B), \quad \text{for all } A, B \in K(X), \tag{8.8}$$

$s_j \in [0, 1)$, $j = 1, \dots, m$.

It can be easily checked that such an IFS consists of n^m contractions with factors at most $s = \prod_{j=1}^m s_j$. By the same arguments as above, there exists a unique fixed point $B^* \in K(X)$ of the composition $\mathscr{F}_m \circ \cdots \circ \mathscr{F}_1$, i.e. $B^* = \mathscr{F}_m \circ \cdots \circ \mathscr{F}_1(B^*)$, such that

$$\lim_{k \to \infty} d_H\left(B^*, (\mathscr{F}_m \circ \cdots \circ \mathscr{F}_1)^k(B_0)\right) = 0, \quad \text{for every } B_0 \in K(X), \qquad (8.9)$$

and

$$d_H\left(B^*, (\mathscr{F}_m \circ \cdots \circ \mathscr{F}_1)^k(B_0)\right) \leq \frac{s^k}{1-s} d_H\left((\mathscr{F}_m \circ \cdots \circ \mathscr{F}_1)^k(B_0), B_0\right), \quad (8.10)$$

for $k = 1, 2, \ldots$

Moreover, its similarity dimension $\dim B^*$ can be estimated from above (in the particular cases even calculated) by means of the formula (cf. (8.6))

$$D = \frac{m \log n}{\log \frac{1}{s}}, \quad s = \prod_{j=1}^m s_j, \qquad (8.11)$$

provided the appropriate open set condition is satisfied.

Corollary 8.1. *For the IFS, associated with the composition*

$$\mathscr{F}_m \circ \cdots \circ \mathscr{F}_1 \colon K(X) \to K(X),$$

where the Hutchinson–Barnsley operators $\mathscr{F}_j$, $j = 1, \ldots, m$, are defined in (8.7) and satisfy (8.8), there exists a unique compact invariant set $B^* \subset X$, i.e. $B^* = \mathscr{F}_m \circ \cdots \circ \mathscr{F}_1(B^*)$, which satisfies the relations (8.9), (8.10). Its fractal (similarity) dimension $D = \dim B^*$ can be calculated by means of formula (8.11), provided (cf. (8.7), (8.8))

$$d(f_{i,j}(x), f_{i,j}(y)) = s_j d(x,y), \quad \text{for all } x, y \in X, \qquad (8.12)$$

where $s_j \in [0,1)$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, and the appropriate open set condition is satisfied.

8.3   Hurst exponent vs. fractal dimension

As indicated in Chapter 4 (cf. also Chapter 2), the Moran–Hutchinson formula (8.5), resp. its particular form (8.6), can be interpreted in terms of the *Zipf–Mandelbrot law* and the *Pareto law* in economics, and in terms of the Menzerath–Altmann law in quantitative linguistics. On the other hand, the authors usually assume strictly self-similar structures in their studies, which is neither natural nor suitable for practical applications. In order to avoid this handicap, we require that the structures are self-similar "only" cyclically (in blocks) like in Corollary 8.1.

As we will show, we can also use an isomorphism between the block-like version (8.11) of the Moran–Hutchinson formula (8.6) and the Hurst formula, originally derived in hydrology (cf. Hurst, 1951; Mosteller, Wallace, 1964), and successfully applied in stock markets (see e.g. Bianchi, Pianese, 2018; Mandelbrot,

2007) or in quantitative linguistics (see e.g. Ausloos, 2012a, 2012b; Hřebíček, 1998a, 2000).

The *(asymptotic) Hurst formula* takes the form (see e.g. Feder, 1988; Hurst, 1951; Hurst, Black, Simaika, 1965; Mosteller, Wallace, 1964)

$$E\left[\frac{R(\tau)}{S(\tau)}\right] = C\tau^{H}, \quad \text{as } \tau \to \infty, \ (C \geq 0 \text{ is a constant}), \qquad (8.13)$$

where the constant $H \geq 0$ is called the *Hurst* (sometimes also *Hurst–Hölder*) *exponent*, which is usually referred as the "index of dependence" or (because of an asymptotic character of (8.13)) as the "index of long-range dependence" of the expected value $E\left[\frac{R(\tau)}{S(\tau)}\right]$ of $\frac{R(\tau)}{S(\tau)}$ on the increasing number τ of data points in a time series. For a partial time series of length τ, namely $x_1, x_2, \ldots, x_\tau$, $R(\tau)$ denotes the range of the first τ cumulative deviations from the mean

$$\overline{x}(\tau) = \frac{1}{\tau} \sum_{i=1}^{\tau} x_i,$$

i.e.

$$R(\tau) := \max_{1 \leq t \leq \tau} x(t, \tau) - \min_{1 \leq t \leq \tau} x(t, \tau),$$

where

$$x(t, \tau) := \sum_{i=1}^{t} (x_i - \overline{x}(\tau)), \quad t = 1, 2, \ldots, \tau,$$

and $S(\tau)$ denotes their standard deviation, i.e.

$$S(\tau) := \sqrt{\frac{1}{\tau} \sum_{i=1}^{\tau} (x_i - \overline{x}(\tau))^2}.$$

The expression $R(\tau)/S(\tau)$ so denotes the rescaled range depending on the time span.

For a *"finite time" Hurst formula*

$$\frac{R(\tau)}{S(\tau)} = C_\tau \tau^{H_\tau}, \quad \text{where } S(\tau) > 0, \qquad (8.14)$$

the related Hurst exponent $H_\tau = H(\tau)$ can be easily calculated as follows ($\tau > 1$):

$$H_\tau = \frac{\log \frac{1}{C_\tau} \frac{R(\tau)}{S(\tau)}}{\log \tau}. \qquad (8.15)$$

If for a sufficiently large values of τ (if they exist) the values H_τ and C_τ, do not change much (i.e. if they do not differ much each from other), then they can be regarded as a suitable approximation of the Hurst exponent H and constant C. Geometrically, it means that plotting $\log\left(\frac{1}{C_\tau} \frac{R(\tau)}{S(\tau)}\right)$ as a function of $\log \tau$, its graph should fit well a straight line whose slope approximates H.

If the values H_τ in (8.15) obtained in this way belong for large τ to the interval $(0.5, 1]$, then a time series exhibits a long-term positive autocorrelation. Otherwise, for $H_\tau \in [0, 0.5)$, they exhibit a negative autocorrelation (erratical oscillation). For $H_\tau = 0.5$, there is no correlation in a time series, and the related increments $x_{t_2} - x_{t_1}$ and $x_{t_3} - x_{t_2}$ are independent in the sense of probability theory.

It is also well known (see e.g. Mandelbrot, 1985; Peitgen, Jürgens, Saupe, 2004, Chapter 9.5) that for self-similar time series, describing a "one-dimensional" random process, the fractal (box-counting) dimension D of their graphs satisfies the equality

$$D = 2 - H, \quad \text{where } D \in [1, 2]. \tag{8.16}$$

Nevertheless, as pointed out in Peitgen, Jürgens, Saupe (2004), Chapter 9.5, the box-counting dimension D in (8.16) may depend on the choice of scales. Moreover, for more general (e.g. a self-affine time series or a multidimensional process), the Hurst exponent H and the fractal dimension D can be chosen rather independently.

The Hurst exponent H represents structures over asymptotically longer periods (requiring in Mandelbrot's terms (cf. Mandelbrot, 1985), the global extrapolative procedures), while fractal dimension D represents structures over asymptotically shorter periods (requiring in Mandelbrot's terms (cf. Mandelbrot, 1985), the local interpolative procedures). Despite it, Mandelbrot was able to show even for more general self-affine structures that the local value of so called *"divider"* dimension D satisfies the equality

$$D = \frac{1}{H}. \tag{8.17}$$

This equality stimulated us to use an isomorphism between (8.6) and (8.15), resp. its reciprocal form

$$\frac{1}{H} = \frac{\log \tau}{\log\left(\frac{1}{C} \cdot \frac{R}{S}\right)}, \tag{8.18}$$

i.e.

$$D \sim \frac{1}{H}, \quad n \sim \tau, \quad r \sim C\frac{S}{R}.$$

More generally, for a cyclically repeated block of time series, characterized by the following Hurst formulas on m scaling levels

$$\frac{R_j}{S_j} = C_j \tau_j^{H_j}, \quad j = 1, \ldots, m, \tag{8.19}$$

resp.

$$H_j = \frac{\log \frac{1}{C_j} \cdot \frac{R_j}{S_j}}{\log \tau_j} = \frac{\log\left(\frac{1}{C_j} \cdot \frac{R_j}{S_j}\right)^k}{\log \tau_j^k}, \quad k \in \mathbb{N}, \; j = 1, \ldots, m, \tag{8.20}$$

we can use the correspondence (cf. formulas (8.11), (8.17) and (8.20)):

$$D \sim \frac{1}{H}, \quad n \sim \tau^k := \tau_1^k = \cdots = \tau_m^k, \quad s_j \sim \left(C_j \frac{S_j}{R_j}\right)^k \in [0, 1), \quad j = 1, \ldots, m, \tag{8.21}$$

where k is a sufficiently big positive integer, in order to formulate an analogy of Corollary 8.1 in terms of the Hurst exponents.

By the *cyclicity hypothesis* we understand that

$$\tau_i = \tau_{i+m}, \quad \frac{R_i}{S_i} = \frac{R_{i+m}}{S_{i+m}}, \quad C_i = C_{i+m}, \quad H_i = H_{i+m}, \quad i = 1, \dots \qquad (8.22)$$

Proposition 8.2. (cf. Andres, Langer, Matlach, 2020). *For any cyclically repeated family of time series, satisfying formulas* (8.19), (8.20) *and* (8.22), *we can associate, via the correspondence* (8.21), *an iterated function system consisting of τ^{km} contractions, which are the m-tupled compositions of functions*

$$f_{i,j} \colon \mathbb{R}^k \to \mathbb{R}^k, \quad i = 1, \dots, \tau^k, \ j = 1, \dots, m.$$

If these functions satisfy the similarity condition (cf. (8.12))

$$\left| f_{i,j}(x) - f_{i,j}(y) \right| = \left(C_j \frac{S_j}{R_j} \right)^k |x - y|, \quad \text{for all } x, y \in \mathbb{R}^k, \qquad (8.23)$$

where

$$C_i \frac{S_j}{R_j} \in [0, 1), \quad i = 1, \dots, \tau^k, \ j = 1, \dots, m,$$

then there exists a unique compact invariant set $B^* \subset \mathbb{R}^k$ of the composition $\mathscr{F}_m \circ \cdots \circ \mathscr{F}_1$ of m Hutchinson–Barnsley operators $\mathscr{F}_j \colon K(\mathbb{R}^k) \to K(\mathbb{R}^k)$, where

$$\mathscr{F}_j(A) := \bigcup_{i=1}^{\tau^k} f_{i,j}^*(A), \quad A \in K(\mathbb{R}^k), \ j = 1, \dots, m,$$

i.e. $B^* = \mathscr{F}_m \circ \cdots \circ \mathscr{F}_1(B^*)$ which satisfies for $X = \mathbb{R}^k$ and

$$s = \prod_{j=1}^{m} \left(C_j \frac{S_j}{R_j} \right)^k$$

the relations (8.9), (8.10).

Moreover, its Hausdorff dimension $D = \dim B^*$ can be calculated by means of the formula

$$D = \frac{m}{\sum_{j=1}^{m} H_j}, \qquad (8.24)$$

where H_j, $j = 1, \dots, m$, are defined in (8.20), *provided k is a sufficiently big positive integer.*

Remark 8.3. If, for instance, $B^* \subset \mathbb{R}^k$ is a Cartesian product of Cantor sets, then k can be taken as $k \geq \max_{j=1,\dots,m} \frac{1}{H_j}$, in order the appropriate open set condition to be satisfied (see Andres, Rypka, 2012).

Remark 8.4. Observe that, unlike the box-counting dimension $D \in [1,2]$, where $H \in [0,1]$ was defined in (8.14), resp. (8.18), the Hausdorff dimension D in (8.24) can theoretivally exceed higher values than 2, e.g. $D > k-1$, which is certainly due to an open set condition. For a von Koch-like curve of dimension $D = 1.95$ and its open set condition, see Figures 8.1 and 8.2.
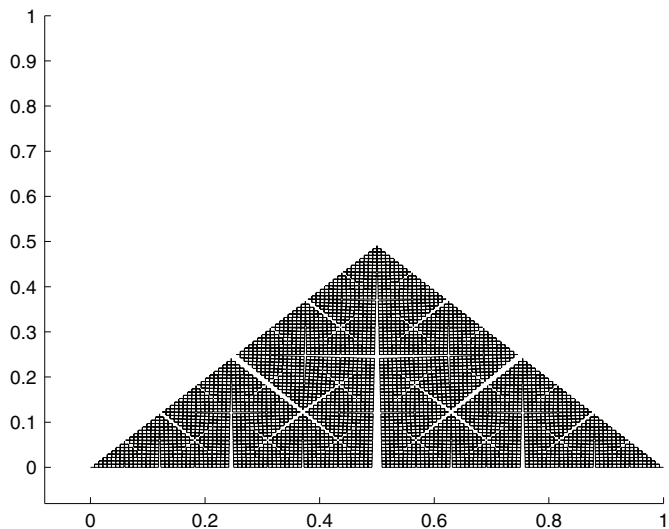

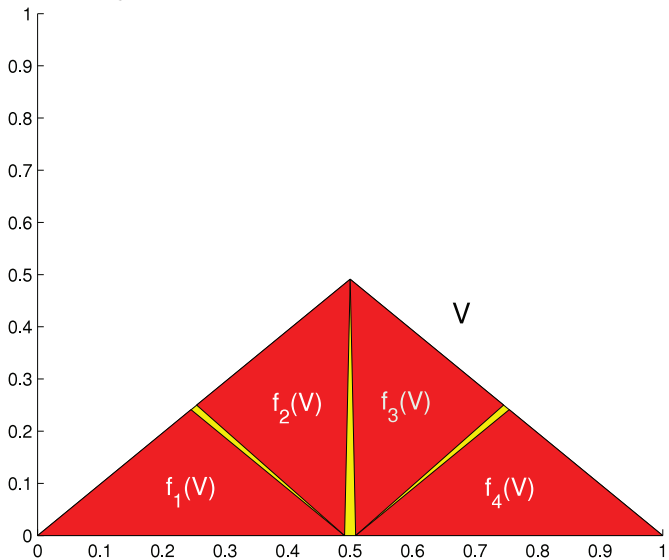Figure 8.1: von Koch-like curve of dimension 1.95

Figure 8.2: Open set for the von Koch-like curve from Fig. 8.1 and its images

8.4   Calculation of the Hurst exponent

In order to calculate the Hurst exponent H for the given sequence $x_1, x_2, \ldots, x_\tau,$
$\ldots, x_N$, where $x_1 \neq x_2$, we can use another isomorphism between (cf. (8.14))

$$\frac{R(\tau)}{S(\tau)} = C\tau^H, \quad \tau = 2, \ldots, N, \tag{8.25}$$

and the (truncated) formula for the *Menzerath–Altmann law*

$$y(x) = Ax^{-b}, \quad x = 1, \ldots, N, \tag{8.26}$$

where $x \in \mathbb{N}$ is the length of a *construct*, $y \in \mathbb{R}$ is the length of a *constituent*,
$A > 0$, b are real constants. For the definition of these notions and more details,
see Chapter 3.

The constants C, H in (8.25) can be calculated in a standard way by the or-
dinary least square regression, including the statistical evaluation, for calculating
the constants A, b in (8.26). On this basis, we can use the correspondence between
(8.25) and (8.26), i.e.

$$\tau \sim x, \quad \frac{R}{S} \sim y, \quad C \sim A, \quad H \sim -b, \tag{8.27}$$

when taking $R(1)/S(1) := 1$.

Thus, we obtain the approximative formulas (cf. Chapter 3):

$$H \approx \tilde{H} := \frac{\sum_{\tau=2}^{N}(\ln \tau)\left(\ln \frac{R(\tau)}{S(\tau)}\right) - \frac{1}{N}\sum_{\tau=2}^{N}\ln \tau \sum_{\tau=2}^{N}\frac{R(\tau)}{S(\tau)}}{\sum_{\tau=2}^{N}(\ln \tau)^2 - \frac{1}{N}\left(\sum_{\tau=2}^{N}\ln \tau\right)^2}, \tag{8.28}$$

$$C \approx \tilde{C} := \exp\left(\frac{1}{N}\sum_{\tau=2}^{N}\ln \frac{R(\tau)}{S(\tau)} - \frac{\tilde{H}}{N}\sum_{\tau=2}^{N}\ln \tau\right). \tag{8.29}$$

The application of these approximative formulas should be evaluated statisti-
cally by means of the following standard criteria (see Chapter 6):
- The *root mean square error* (RMSE).
- The *coefficient of determination* $R^2 \in [0, 1]$.
- *Homoscedasticity*.
- *Normality*.

If we are able to detect m scaling levels, then the constants N, C, $\tilde{C}$, H, $\tilde{H}$ in
(8.28), (8.29) can be appropriately indexed as $N_1, \ldots, N_m$; $C_1, \ldots, C_m$; $\tilde{C}_1, \ldots, \tilde{C}_m$;
$H_1, \ldots, H_m$; $\tilde{H}_1, \ldots, \tilde{H}_m$.

Thus,

$$C_j(\tau)\frac{S_j(\tau)}{R_j(\tau)} \approx \tilde{C}_j\frac{S_j(\tau)}{R_j(\tau)} \approx \tau^{-\tilde{H}_j} \in (0, 1), \tag{8.30}$$

for $\tau = 2, \ldots, N_j$, $j = 1, \ldots, m$, when $\tilde{H}_j > 0$, $j = 1, \ldots, m$.

We are ready to formulate the main theorem.

Theorem 8.1. (cf. Andres, Langer, Matlach, 2020). *Consider the family of m sequences of the form*

$$x_1^{(j)}, x_2^{(j)}, \ldots, x_\tau^{(j)}, \ldots, x_{N_j}^{(j)}, \quad j = 1, \ldots, m,$$

and suppose that (cyclicity hypothesis)

$$x_\tau^{(j)} = x_\tau^{(j+m)}, \quad \text{for all } \tau = 1, \ldots, N_j, \ j = 1, \ldots, m.$$

Assume, furthermore, that the related calculations of constants C_j, H_j, $j = 1, \ldots, m$, by means of approximative formulas (8.28), (8.29) can be evaluated statistically in a positive way (the standard criteria are: RMSE, NRMSE, R^2, homoscedeasticity, normality). Then the Hausdorff dimension $D = \dim B^*$ of the associated, via (8.21), (8.22) (by the procedure described in Proposition 8.2), fractal $B^* \subset \mathbb{R}^k$, where k is a sufficiently big positive integer, equals approximately the value (cf. (8.24)) $m \big/ \sum_{j=1}^m \tilde{H}_j$, i.e.

$$D \approx \frac{m}{\sum_{j=1}^m \tilde{H}_j},$$

provided $\tilde{H}_j > 0$, $j = 1, \ldots, m$.

For the sake of brevity, let us introduce the following definition.

Definition 8.1. The fractal $B \subset \mathbb{R}^k$ in Theorem 8.1 and Proposition 8.2 will be called a *fractal model* of the given family of m sequences, satisfying the cyclicity hypothesis. The arithmetic mean of the partial Hurst exponents H_j, $j = 1, \ldots, m$, will be denoted again as H, i.e.

$$H = \frac{1}{m} \sum_{j=1}^m H_j,$$

and called the *Hurst exponent* of the given family of sequences. Its approximation

$$\tilde{H} := \frac{1}{m} \sum_{j=1}^m \tilde{H}_j$$

(see (8.28)) will be called the *approximative Hurst exponent*.

In view of Definition 8.1, we can immediately reformulate Theorem 8.1 as follows.

Theorem 8.2. *Under the assumptions of Theorem 8.1, the Hausdorff dimension D of the fractal model of the given family of m sequences, satisfying the cyclicity hypothesis, equals approximately the reciprocal value of its approximative Hurst exponent $\tilde{H}$, i.e.*

$$D = \frac{1}{H} \approx \frac{1}{\tilde{H}}, \quad \text{where } \tilde{H} = \frac{1}{m} \sum_{j=1}^m \tilde{H}_j \qquad \text{(see (8.28))},$$

provided $\tilde{H}_j > 0$, $j = 1, \ldots, m$.

Remark 8.5. Although the statistical verification of the calculating procedures in Theorems 8.1, 8.2 are only required for the partial approximative Hurst exponents $\tilde{H}_j$, $j = 1, \ldots, m$, a finite additive (resp. arithmetic averaging) character of

$$\tilde{H} = \frac{1}{m} \sum_{j=1}^{m} \tilde{H}_j$$

allows us to accept its statistical verification, too. Nevertheless, to be precise, one should use the correcting algorithms here.

Remark 8.6. The results obtained by means of Theorems 8.1, 8.2 must be, under the cyclicity hypothesis, interpreted in an exclusively restrictive way up to the given m scaling levels. Despite this restriction, such a situation is much more natural than the strict self-similarity on each scaling level. In the nature, it reminds a finite multiple branching of trees.

8.5   Specifica of quantitative linguistics for sign language

In view of the results like those in Ausloos (2012a, 2012b), Hřebíček (1998a, 2000), the Hurst formula (see Hurst, 1951; Hurst, Black, Simaika, 1965; Mosteller, Wallace, 1964) can aspirate to become a linguistic law, too. We would like to demonstrate here that the same is true for a sign language.

Thus, we can again detect at least four sign language units, namely sentences, clauses, signs and morphemes, on the appropriate sign language (scaling) levels. Their lengths can be measured as follows:
   – in the number of clauses for sentences,
   – in the number of signs for clauses,
   – in seconds for signs.
   Alternatively, they can be measured either all, or some further, in seconds.

In the previous Chapter 7, we were able to detect in this way, by means of the Menzerath–Altmann law, a "sign language fractal of the second order" and to measure its *degree of semanticity*. The notion of "second order" refers to just two respected sign language levels (sentences and clauses). On the lower level of signs, the Menzerath–Altmann law has not been satisfied. Because of the simultaneity of signs, it was complicated and ineffective to measure appropriately their length in the number of morphemes (gestures).

In our experiment, the research data of the Czech sign language are the same as in Chapter 7.

The detection of morphemes (gestures), as the simultaneous constituents of signs, and the measurement of the length of signs in the somehow suitably "elaborated" number of morphemes would be certainly interesting, but because of our negative experience in Chapter 7, we decided to omit it here. For the sake of simplicity, we also avoided to involve into our research all the mimics and non-natural factors.

Our experience in earlier experiments with natural languages texts, investigated by virtue of the Menzerath–Altmann law, indicates that the fractal dimension

of the respective models (we call them *language fractals*) can be interpreted as a *measure of semanticity* of the given text (see Chapter 4). The same might be true for a sign language.

Because of a hydrological analogy (cf. Hurst, 1951; Mosteller, Wallace, 1964), the Hurst exponents might determine the dynamics of "semantic lakes" in our minds.

8.6   Application to sign language texts

Example 8.1. Our experiment concerns again the fractal analysis of a concrete sign language text, considered on three scaling levels:

   I) level of sentences,
   II) level of clauses,
   III) level of signs.

In other words, we will apply Theorems 8.1, 8.2 to the following family of three sequences of sentence, clause and sign lengths:

$$x_1^I = 1, x_2^I = 2, x_3^I = 2, \ldots, x_\tau^I, \ldots, x_{72}^I = 2, x_{73}^I = 4, x_{74}^I = 2,$$

$$x_1^{II} = 5, x_2^{II} = 2, x_3^{II} = 4, \ldots, x_\tau^{II}, \ldots, x_{245}^{II} = 2, x_{246}^{II} = 1, x_{247}^{II} = 1,$$

$$x_1^{III} = 1.278, x_2^{III} = 0.948, x_3^{III} = 1.380, \ldots, x_\tau^{III}, \ldots, x_{891}^{III} = 1.070,$$

$$x_{892}^{III} = 0.434, x_{893}^{III} = 0.223,$$

whose respective numbers of terms are $N_I = 74$, $N_{II} = 247$, $N_{III} = 893$.

The length of a sentence is measured in the number of clauses, the length of a clause is measured in the number of signs and the length of a sign is measured in seconds.

Calculating, by means of the formulas in Section 8.3, the values of the rescaled ranges

$$\frac{R_I(\tau)}{S_I(\tau)}, \quad \tau = 2, \ldots, 74, \tag{8.31}$$

$$\frac{R_{II}(\tau)}{S_{II}(\tau)}, \quad \tau = 2, \ldots, 247, \tag{8.32}$$

$$\frac{R_{III}(\tau)}{S_{III}(\tau)}, \quad \tau = 2, \ldots, 893, \tag{8.33}$$

and applying appropriately the formulas (8.28), (8.29), we obtained in Tables 8.1–8.3, the results for the constants $\tilde{C}_I$, $\tilde{H}_I$, $\tilde{C}_{II}$, $\tilde{H}_{II}$, $\tilde{C}_{III}$, $\tilde{H}_{III}$, including their statistical evaluation.

Table 8.1: Experimental results on the level of sentences

$\tilde{C}_I$	$\frac{1}{\tilde{C}_I}$	$\tilde{H}_I$	RMSE	NRMSE	R^2	Homosc.	Normality
0.683	1.464	0.665	0.152	0.057	0.940	not reject.	not reject.

Table 8.2: Experimental results on the level of clauses

$\tilde{C}_I$	$\frac{1}{\tilde{C}_I}$	$\tilde{H}_I$	RMSE	NRMSE	R^2	Homosc.	Normality
0.642	1.557	0.711	0.102	0.029	0.978	rejected	rejected

Table 8.3: Experimental results on the level of signs

$\tilde{C}_I$	$\frac{1}{\tilde{C}_I}$	$\tilde{H}_I$	RMSE	NRMSE	R^2	Homosc.	Normality
1.633	0.612	0.435	0.188	0.050	0.838	rejected	rejected

The graphs illustrating the dependence of the functions (cf. (8.25))

$$\tilde{C}_I \tau^{\tilde{H}_I}, \quad \tilde{C}_{II} \tau^{\tilde{H}_{II}}, \quad \tilde{C}_{III} \tau^{\tilde{H}_{III}}$$

on τ, calculated from the input data in Tables 8.1–8.3, are plotted (jointly with the values of rescaled ranges) in Figures 8.3–8.5. It will be also convenient to present the graphs illustrating the dependence of the reciprocal functions

$$\frac{1}{\tilde{C}_I} \tau^{-\tilde{H}_I}, \quad \frac{1}{\tilde{C}_{II}} \tau^{-\tilde{H}_{II}}, \quad \frac{1}{\tilde{C}_{III}} \tau^{-\tilde{H}_{III}}$$

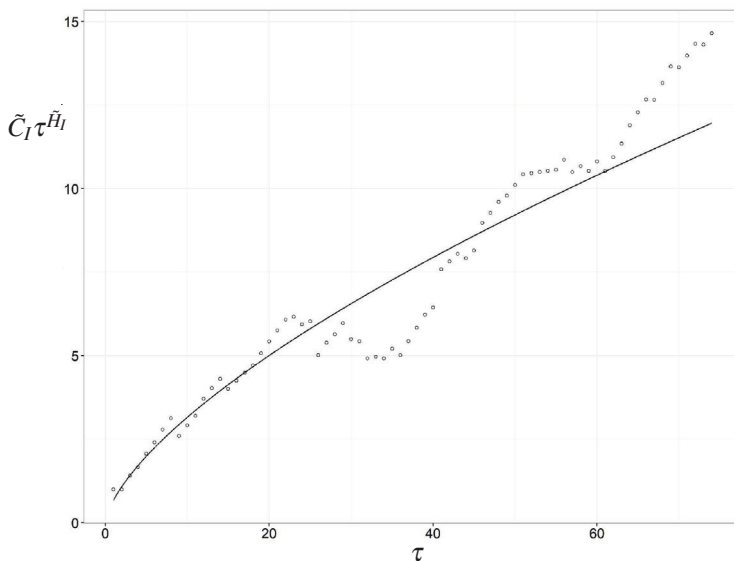on τ (jointly with the reciprocal values of rescaled ranges) in Figures 8.6–8.8.

Figure 8.3: The graph of $\tilde{C}_I \tau^{\tilde{H}_I}$ vs. τ calculated from the input data in Table 8.1. The dots indicate the values of the rescaled ranges $\frac{R_I(\tau)}{S_I(\tau)}$.
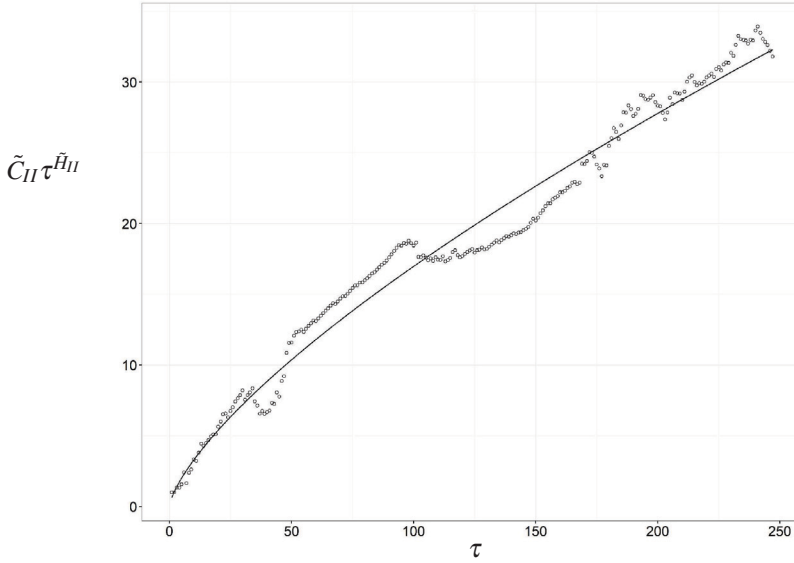
Figure 8.4: The graph of $\tilde{C}_{II}\tau^{\tilde{H}_{II}}$ vs. τ calculated from the input data in Table 8.2. The dots indicate the values of the rescaled ranges $\frac{R_{II}(\tau)}{S_{II}(\tau)}$.
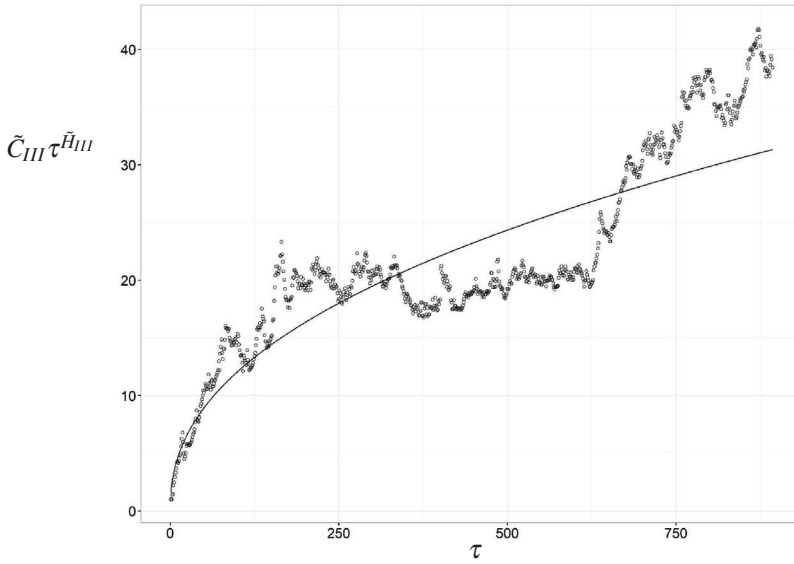

Figure 8.5: The graph of $\tilde{C}_{III}\tau^{\tilde{H}_{III}}$ vs. τ calculated from the input data in Table 8.3. The dots indicate the values of the rescaled ranges $\frac{R_{III}(\tau)}{S_{III}(\tau)}$.
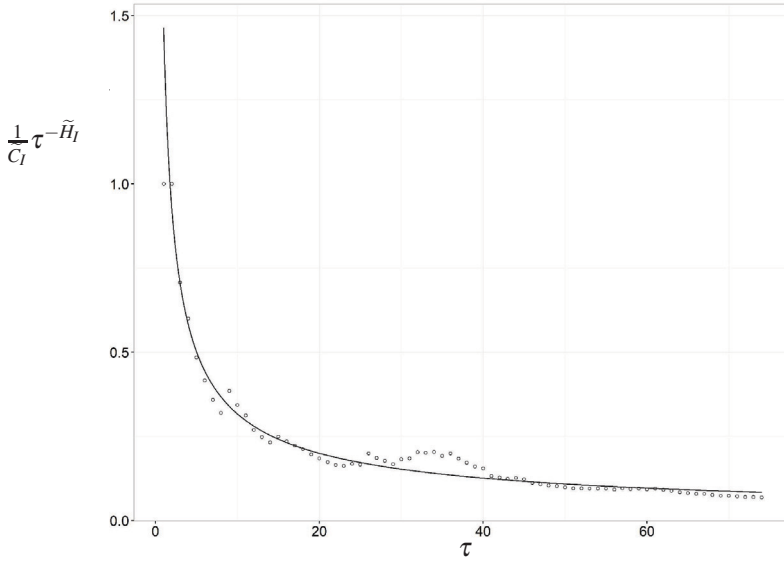
Figure 8.6: The graph of $\frac{1}{C_I}\tau^{-\tilde{H}_I}$ vs. τ calculated from the input data in Table 8.1. The dots indicate the values of $\frac{S_I(\tau)}{R_I(\tau)}$.
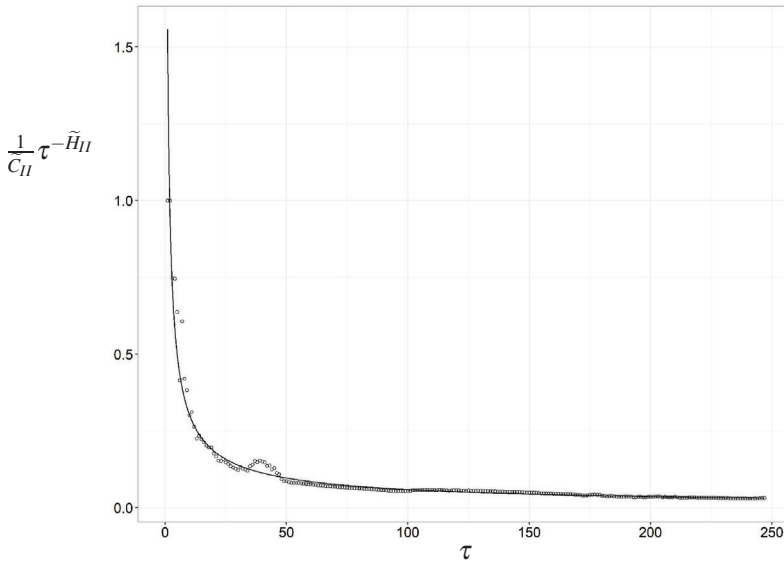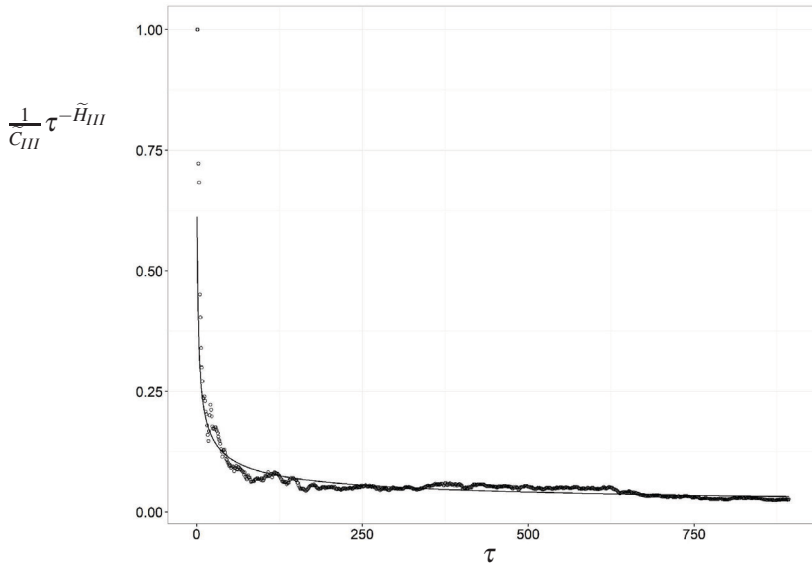

Figure 8.7: The graph of $\frac{1}{C_{II}}\tau^{-\tilde{H}_{II}}$ vs. τ calculated from the input data in Table 8.2. The dots indicate the values of $\frac{S_{II}(\tau)}{R_{II}(\tau)}$.

Figure 8.8: The graph of $\frac{1}{\tilde{C}_{III}}\tau^{-\tilde{H}_{III}}$ vs. τ calculated from the input data in Table 8.3. The dots indicate the values of $\frac{S_{III}(\tau)}{R_{III}(\tau)}$.

Since the calculated values of Hurst exponents $\tilde{H}_I, \tilde{H}_{II}, \tilde{H}_{III}$ are obviously positive and the calculations were statistically evaluated in a positive way, we can apply Theorem 8.2 for the following conclusion.

Conclusion 8.3. *The Hausdorff dimension D of the fractal model of the family of three sequences, associated with a given sign language text (speech), satisfies under the cyclicity hypothesis $D \doteq 1.656\,668$. For the approximative Hurst exponent $\tilde{H}$, we obtained that*

$$\tilde{H} = \frac{1}{3}\left(\tilde{H}_I + \tilde{H}_{II} + \tilde{H}_{III}\right) = 0.603\,621,$$

which is not far from the values around 0.73, usually observed in natural processes (see e.g. Hřebíček, 2000; Hurst, Black, Simaika, 1965).

According to Theorem 8.2, we have namely that

$$D = \frac{1}{H} \approx \frac{1}{\tilde{H}} = \frac{3}{\tilde{H}_I + \tilde{H}_{II} + \tilde{H}_{III}} = 1.656\,668.$$

8.7   Concluding discussion

Observe that the numerical calculations in Example 8.1 have been statistically justified, especially because of sufficiently law values of the normalized root mean square errors (NRMSE) and high values of the coefficient of determination R^2 in Tables 8.1–8.3.

On the other hand, since the normality of given data sets was, except sentences, rejected, on two sign language (scaling) levels under our consideration, i.e. that their distribution was not there Gaussian, the respective confidence intervals for the calculated values of constants $\tilde{C}_I, \tilde{C}_{II}, \tilde{C}_{III}$ and $\tilde{H}_I, \tilde{H}_{II}, \tilde{H}_{III}$ could not be determined in a standard way by the Wald statistics. The preferable appropriate way of their determination would be by bootstrapping (see e.g. Efron, Tibshirani, 1993, Weron, 2002), but to avoid the technicalities, we decided to omit it here.

The dispersion of random errors, related to a process of approximation, occured to be constant (i.e. homoscedasticity) again only on the level of sentences.

By the same reasons, we decided to omit the negligible corrections here, too.

Since

$$\tilde{C}_I \frac{S_I(\tau)}{R_I(\tau)} = 0.682\,849 \frac{S_I(\tau)}{R_I(\tau)} \leq 0.682\,849,$$

$$\tau^{-\tilde{H}_I} \leq 2^{-\tilde{H}_I} = 2^{-0.665\,168} = 0.630\,615$$

hold for all $\tau = 2, \ldots, 74$, and

$$\tilde{C}_{II} \frac{S_{II}(\tau)}{R_{II}(\tau)} = 0.642\,424 \frac{S_{II}(\tau)}{R_{II}(\tau)} \leq 0.642\,424,$$

$$\tau^{-\tilde{H}_{II}} \leq 2^{-\tilde{H}_{II}} = 2^{-0.710\,961} = 0.610\,913$$

hold for all $\tau = 2, \ldots, 247$, conditions (8.30) (cf. (8.23)) is satisfied on the scaling levels of sentences and clauses.

On the other hand, although

$$\tau^{-\tilde{H}_{III}} \leq 2^{-\tilde{H}_{III}} = 2^{-0.434\,735} = 0.739\,830$$

holds for all $\tau = 2, \ldots, 893$, the relations

$$\tilde{C}_{III} \frac{S_{III}(\tau)}{R_{III}(\tau)} = 1.633\,110 \frac{S_{III}(\tau)}{R_{III}(\tau)} \leq 0.736\,137$$

hold only for $\tau = 5, \ldots, 893$. For $\tau = 2, 3, 4$, we have

$$1.633\,110 \frac{S_{III}(\tau)}{R_{III}(\tau)} > 1.$$

Condition (8.30) (cf. (8.23)) is therefore satisfied, on the scaling level of signs, only for $\tau = 5, \ldots, 893$.

Fortunately,

$$\tilde{C}_I \frac{S_I(\tau)}{R_I(\tau)} \cdot \tilde{C}_{II} \frac{S_{II}(\tau)}{R_{II}(\tau)} \cdot \tilde{C}_{III} \frac{S_{III}(\tau)}{R_{III}(\tau)} \leq \tilde{C}_I \tilde{C}_{II} \tilde{C}_{III} = 0.716\,410$$

holds for all $\tau = 2, 3, 4$, which is sufficient for satisfying the contraction condition in the proof of Proposition 8.2, and subsequently of Theorems 8.1, 8.2.

In view of this slight modification, it allows us to write that, under the cyclicity hypothesis,

$$D = \frac{1}{H} \approx \frac{1}{\tilde{H}} = \frac{3}{\tilde{H}_I + \tilde{H}_{II} + \tilde{H}_{III}} = 1.656668$$

holds for the Hausdorff dimension D of the fractal model of the family of three sequences, associated with a given sigh language text (speech), as claimed. Its box-counting dimension satisfies approximately

$$2 - \tilde{H} = 2 - \frac{1}{3}\left(\tilde{H}_I + \tilde{H}_{II} + \tilde{H}_{III}\right) = 2 - 0.603621 = 1.396379,$$

jointly with

$$2 - \tilde{H}_I = 2 - 0.665168 = 1.334832,$$
$$2 - \tilde{H}_{II} = 2 - 0.710961 = 1.289039,$$
$$2 - \tilde{H}_{III} = 2 - 0.434735 = 1.565265.$$

In both cases, the values indicate some volatility, roughness and a not completely smooth trend. The appropriate fractal model can be constructed in a von Koch-like way (see Andres, Rypka, 2012, Theorem 2).

Since the value of the approximative Hurst exponent $\tilde{H}$ is higher than 0.5, namely

$$\tilde{H} = \frac{1}{3}\left(\tilde{H}_I + \tilde{H}_{II} + \tilde{H}_{III}\right) = 0.603621,$$

where $\tilde{H}_I = 0.665168$, $\tilde{H}_{II} = 0.710961$, $\tilde{H}_{III} = 0.434735$, the process as a whole is persistent (non-Markovian) with some positive memory. On the other hand, it is persistent on the levels of sentences and clauses, but not on the lowest level of signs, where $\tilde{H}_{III} = 0.434735$ and $1/\tilde{H}_{III} = 2.300252$ (see Remark 8.4).

Peter Angermann, *A. I. (Nohandsfractal)*, 2010, crayon

9   SOME FURTHER POSSIBILITIES

This book is rather a collection of interdisciplinary ingredients than a systematically written monograph. On the other hand, as far as we know, it is the first attempt in the field of a quantitative linguistic approach to a sign language. As every pioneering work, it suffers from many deficiencies.

The signs of the deaf can be regarded as analogies to words. On the other hand, the relationship between the respective morphemes seems to be, in view of their simultaneous (multidimensional) character, quite sophisticated and unclear. One must therefore especially overcome the difficulties related to a simultaneous action of morphemes, as the constituents of signs. It is a question whether or not some "linearization" or "averaging" technique can help us to measure appropriately the lengths of signs in "suitable quantities" of morphemes, or so. For more details about a simultaneous (nonlinear) character of sign languages, see e.g. Bellugi, Fischer (1972), Biberauer, Roberts (2013), Hadler, Irrgang (2014), Jammer (2006), Klima, Bellugi (1979), Kremers (2009, 2012), Lillo-Martin, Gajewski (2014), Natsopoulos, Abadzi (1986), Vermeerbergen, Leeson, Crasborn (2007).

It has certainly a meaning to investigate still supra-sentence units of a sign language, analogously to spoken languages, in order the sentences to be their constituents. For spoken languages, such units were discovered by Hřebíček, and named *semantic constructs* (see Hřebíček, 1992, 1995).

In our future research, we would also like to prepare and carry out further experiments about the (fractal-based) frequency analysis of sign language texts (speeches), by means of the Hurst formula and via the Zipf–Mandelbrot law. The dynamic aspects of sign languages, like (in) stability, predictability and deterministic chaos, can be explored by virtue of Lyapunov exponents, etc.

Despite the mentioned handicaps, we hope that our book can serve at least as an introduction for further steps to this meritorious interdisciplinary area.

Peter Angermann, *A. I. (Nohandsfractal)*, 2016, silkscreen

10 CONCLUSION

The goal of this book was to enrich the theoretical knowledge about Czech sign language by means of quantitative linguistic analyses and procedures. It was a great challenge, because there has not been any comparative study from such a perspective to our disposal. One of the partial purposes was, besides other things, to draw conclusions usable to the didactics of teaching the Czech sign language.

The key experiment performed to enrich the knowledge of didactics of sign language teaching is a study focused on the subjective perception of the complexity of signs. The partial goal of this study was to verify the proposed method of identifying constituents of signs, which we indicated partly in Chapter 5. A detailed description of the study, including complete results and analyses, is given in the paper Langer, Rypka (2017)

In the professional literature, the signs of sign languages are grouped in several classes according to so-called motion matrices. According to Battison (1978), the following classes are the most common:

- Signs that may be articulated by one hand:
  - signs with zero contacts (e.g. STUDY, DAY, SUN),
  - signs with a contact with the body, however, not with the other hand (e.g. DEAF, FRIEND, OLD).
- Signs articulated by both hands:
  - signs in which both hands are active and both of them are in the same shape (e.g. WEATHER, IMPORTANT, SLOVAKIA);
  - symmetric signs in which one hand is active (the dominant hand—i.e. the right hand among the right-handed people) which articulates above (under, behind, in front of, etc.) the passive hand; both hands are in the same shape (e.g. SOCK, OCTOBER, DANGER); o asymmetric signs in which only one hand is active (the dominant hand) which articulates above (under, behind, in front of, etc.) the passive hand; both hands are in a different shape (e.g. GIFT, TECHNOLOGY, TACK).
- Composed signs which include combinations of signs of types mentioned above (e.g. MAY, LIBRARY, DO NOT KNOW).

For the needs of our research and for exploration of the subjective perception of complexity of signs to be learnt, it was necessary to create a classification of motion matrices, which respects also the didactic difficulty of individual signs:

- Signs articulated by one hand:
  - signs with zero contacts (= 1H without contact),
  - signs with a contact with the body, however, not with the other hand (= 1H with contact).
- Signs articulated by both hands:
  - symmetric signs in which both hands are active and both of them are in the same shape and perform the same (mirror) motion (= 2H symmetric);
  - semi-symmetric in which the hands are in the same shape and they perform similar or phase-shifted motion (= 2H semisymmetric);
  - non-symmetric signs in which the hands are in the same shape, however, only one hand is active (= 2H non-symmetric);
  - asymmetric signs in which the hands are in different shapes and only one hand is active (= 2H asymmetric).
- Composed signs which include combinations of all above mentioned types of signs (= composed).

All seven motion matrices were accompanied by four corresponding signs of Czech sign language among which the sum of morphemes was analysed. After that, they were anonymized and they were subjected to assessment of respondents of the research in random order. Four signs of each motion matrix seemed optimal number with a regard to the variability of summarized number of morphemes, in order to provide a sufficient number of research data and also with regard to adequate time demands for respondents who had to assess 28 signs in total.

The tested signs may be watched (in a form of video recording) at links below:

- 1H without contact—YES (S02), 100 (S05), SUN (S10), WOLF (S18): `https://youtu.be/VYzg_bgutxA`
- 1H with contact—SPAIN (S11), OLD (S12), ART (S14), GREEN (S27): `https://youtu.be/HUdvv7orIHs`
- 2H symmetric—WEATHER (S08), HELP (S22), FAMILY (S24), EGG (S26): `https://youtu.be/kn5Fc8Lblik`
- 2H semisymmetric—DUCK (S04), HIGHWAY (S09), SHOP (S21), CABBAGE (S28): `https://youtu.be/7QLuwpoizkE`
- 2H non-symmetric—BANANA (S01), FAT (S13), FAITHFUL (S17), DANGER (S20): `https://youtu.be/F7Puma2tJVU`
- 2H asymmetric—TACK (S03), TRUST (S16), BRIGHT (S19), YEAR (S25): `https://youtu.be/5Ec0INzhYoc`
- Composed—LIBRARY (S06), MAY (S07), FEBRUARY (S15), DON'T KNOW (S23): `https://youtu.be/Znfe8Ikk770`

In order to discover the subjective perception of signs' complexity, an online questionnaire was applied (via GoogleDocs) while it employed five-degree Likert's scale to mark the complexity of all 28 signs. Individual signs were introduced to the respondents in anonymized form of isolated video recordings at which a deaf signer interprets the sign. All video recordings of signs were presented in the same light and color conditions and in the same (standard) speed. The respondents might (according to the introductory instruction) play the video recording of each sign any time they wanted and they were asked to mark their subjective

perception of each sign on the scale of 1 (very simple) – 5 (very complicated) (see Figure 10.1). It was recommended to watch all the signs at the beginning and to use the whole scale 1–5.

Figure 10.1: On-line form for the scale-based assessment of signs' complexity

Due to a fact that a vast part of Czech sign language is visually motivated, has an iconic nature and, therefore, the visual course of a sign reminds a typical attribute of an item or a phenomenon which they represent (cf. Langer, 2013), it is possible to think that the knowledge of semantic meaning of these signs may influence the respondent in their subjective perception of sign's complexity. Since the aim of the realized research was to discover the perception of signs' complexity of respondents who were not aware if their meaning, a question to discover respondent's practical experience with Czech sign language (or for how long do they study Czech sign language) was included in the questionnaire as well. At the end of the questionnaire, several additional questions of demographic nature were included to discover the sex, age and economic status (high school student, university student, other respondents) of respondents. The respondents might also add any comments to the researched problematics and the online questionnaire itself. In total, 236 unique respondents were joined the research while 1 respondent have not managed to fill in the whole online questionnaire correctly. Therefore, this respondent's answers were not included in the further process. Therefore, data by 235 unique respondents were subjected to further analyses.

From the responses which were recorded in data file of online questionnaire form, the following results of the scale-based assessment of signs complexity (1 = very easy; 5 = very hard) emerged for the researched signs (see Table 10.1):

Table 10.1: Average assessment of signs' complexity by the respondents

Sign (English equivalent)	Anonymous code	Motion matrix	Number of morphemes	Average rating		
				All	Without experience	With experience
YES	S02	1H without contact	6	1.3	1.3	1.3
100	S05	1H without contact	11	3.2	3.1	3.4
SUN	S10	1H without contact	7	1.2	1.3	1.1
WOLF	S18	1H without contact	9	1.5	1.5	1.6
GREEN	S27	1H with contact	7	1.6	1.6	1.7
SPAIN	S11	1H with contact	6	2.2	2.2	2.4
OLD	S12	1H with contact	8	1.4	1.4	1.3
ART	S14	1H with contact	11	2.8	2.9	2.6
WEATHER	S08	2H symmetric	15	1.3	1.3	1.2
HELP	S22	2H symmetric	11	1.3	1.4	1.2
FAMILY	S24	2H symmetric	17	2.0	2.0	1.8
EGG	S26	2H symmetric	19	1.6	1.6	1.4
DUCK	S04	2H semisymmetric	19	3.0	3.0	2.7
HIGHWAY	S09	2H semisymmetric	15	2.0	2.0	1.9
SHOP	S21	2H semisymmetric	13	1.7	1.7	1.6
CABBAGE	S28	2H semisymmetric	17	2.0	2.1	1.9
BANANA	S01	2H non-symmetric	16	2.7	2.9	1.8
FAT	S13	2H non-symmetric	13	1.8	1.7	1.8
FAITHFUL	S17	2H non-symmetric	11	1.4	1.3	1.5
DANGER	S20	2H non-symmetric	14	2.8	2.8	2.6
TACK	S03	2H asymmetric	11	1.6	1.5	2.0
TRUST	S16	2H asymmetric	14	1.8	1.7	2.1
BRIGHT	S19	2H asymmetric	16	3.1	3.1	3.1
YEAR	S25	2H asymmetric	18	2.1	2.2	1.9
LIBRARY	S06	composed	24	3.6	3.7	3.0
MAY	S07	composed	18	3.5	3.4	3.5
FEBRUARY	S15	composed	31	3.1	3.2	2.7
DON'T KNOW	S23	composed	15	3.4	3.4	3.3

From the results stated above, it is possible to conclude (see Table 10.2) that the lowest average values at scale from 1 (very simple) to 5 (very complicated) were achieved in all three groups (all/without any experience with sign language/with experience with sign language) by one-handed signs or by two-handed signs in which both hands are in the same shape (the symmetric and non-symmetric). Signs of this type therefore have an extensive didactic potential and they should be (in addition to the contextual and content point of view) included in the vocabulary for students—beginners.

Table 10.2: The order of seven signs with the lowest average value of complexity

Rank	All			Without experience			With experience		
	Sign (English equivalent)	Motion matrix	Av. rating	Sign (English equivalent)	Motion matrix	Av. rating	Sign (English equivalent)	Motion matrix	Av. rating
1.	SUN	1H without contact	1.2	SUN	1H without contact	1.3	SUN	1H without contact	1.1
2.	WEATHER	2H symmetric	1.3	FAITHFUL	2H non-symmetric	1.3	WEATHER	2H symmetric	1.2
3.	HELP	2H symmetric	1.3	WEATHER	2H symmetric	1.3	HELP	2H symmetric	1.2
4.	YES	1H without contact	1.3	YES	1H without contact	1.3	OLD	1H with contact	1.3
5.	FAITHFUL	2H non-symmetric	1.4	HELP	2H symmetric	1.4	YES	1H without contact	1.3
6.	OLD	1H with contact	1.4	OLD	1H with contact	1.4	EGG	2H symmetric	1.4
7.	WOLF	1H without contact	1.5	WOLF	1H without contact	1.5	FAITHFUL	2H non-symmetric	1.5

In terms of a group of 7 signs with the highest average value of complexity, the vast majority of respondents (as expected) of all 3 groups mentioned the composed signs—i.e. signs with a combination of more motion matrices in which there is a larger number of morphemes (see Table 10.3). However, the placement of a sign for a numeral 100 with a motion matrix of a one-handed sign without a contact is rather surprising. Nevertheless, it is a sign with rather high number of morphemes which are necessary to capture—therefore, it is obvious that the high number of morphemes is significantly projected in the perception of its complexity even in case of one-handed sign. From the perspective of didactics, they are not appropriate to be taught in the initial training.

Table 10.3: The order of seven signs with the highest average value of complexity

Rank	All			Without experience			With experience		
	Sign (English equivalent)	Motion matrix	Av. rating	Sign (English equivalent)	Motion matrix	Av. rating	Sign (English equivalent)	Motion matrix	Av. rating
22.	DUCK	2H semi-symmetric	3.0	DUCK	2H semi-symmetric	3.0	FEBRUARY	composed	2.7
23.	FEBRUARY	composed	3.1	100	1H without contact	3.1	DUCK	2H semi-symmetric	2.7
24.	BRIGHT	2H asym-metric	3.1	BRIGHT	2H asym-metric	3.1	LIBRARY	composed	3.0
25.	100	1H without contact	3.2	FEBRUARY	composed	3.2	BRIGHT	2H asym-metric	3.1
26.	DON'T KNOW	composed	3.4	DON'T KNOW	composed	3.4	DON'T KNOW	composed	3.3
27.	MAY	composed	3.5	MAY	composed	3.4	100	1H without contact	3.4
28.	LIBRARY	composed	3.6	LIBRARY	composed	3.7	MAY	composed	3.5

In Figure 10.2, the correlation matrix for signs and assessments in the questionnaire is plotted. The most interesting discoveries are presented in fourth to

sixth rows. The assessment of complexity done by respondents with and without experience with sign language are strongly correlated. Therefore, the signs which are complex for one group are complex for the second group as well. Among both groups, the correlation is growing when the number of morphemes is adjusted (last two columns), therefore, the adjusted assessment is more corresponding to the perception of complexity of signs among the respondents (the correlation raises from 0.42 to 0.57 among respondents without experience, and from 0.57 to 0.68 in total.

From the didactic perspective, we consider the confirmation of the theoretical assumption of the easier memorability of the signs, which contain a lesser number of morphemes. As the simplest ones, the symmetric signs are considered, i.e. those which are performed with hands in the same shape, they are in mirror positions and they perform the identical motion. In addition, both groups of one-handed signs are perceived as relatively simple—e.g. those with no contact and with a contact to a body. The composed signs are unambiguously perceived the most complex and difficult. While designing the methodology of Czech sign language teaching for target groups of hearing people (while the research included 235 respondents from 13 to 42 years of age), it would be appropriate to include signs from groups of motion matrices which were perceived simple mainly in the initial stages of the teaching.
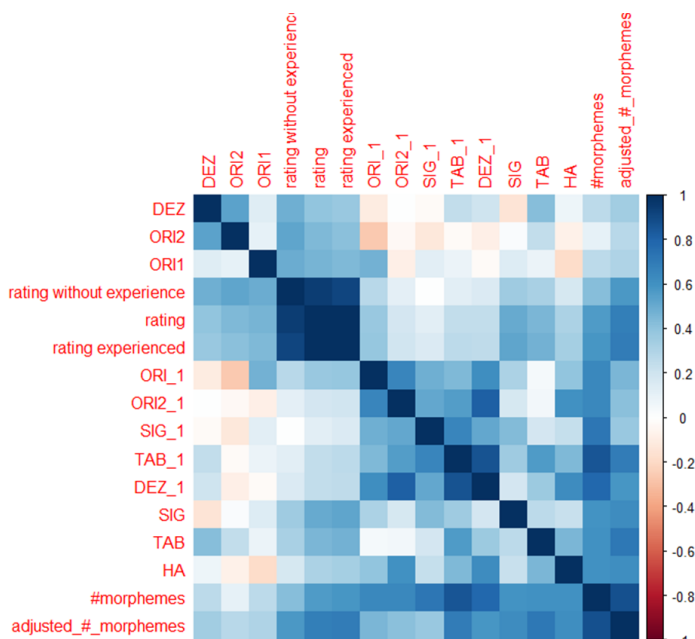

Figure 10.2: Correlation of a type of morpheme and its assessment[2]

[2]TAB = place of articulation; DEZ = shape of the articulating hand(s); ORI1 = orientation of palm; ORI2 = orientation of fingers; HA = mutual position of hands; SIG = hand movements. For detailed desctiption of sign parameters see Section *Manual factors in sign language* in Chapter 1.3 *Sign language as a natural language*

Unfortunately, we have not yet been able so far to put into effect the results of the questionnaire with respect to a possible connection to a methodology how to measure effectively the lengths of signs and their constituents. Nevertheless, the performed experiments about the verification of quantitative linguistic laws for the speeches in Czech sign language opened the new horizons in sign languages investigations. On the other hand, the sign language effects (e.g. synergy of simultaneities), which have no analogy in spoken languages, can also bring new impulses into quantitative linguistics itself. For sentences and clauses, the obtained results hold in analogy to spoken languages.

Peter Angermann, *A. I. (Nohandsfractal)*, 2001, animated drawing

BIBLIOGRAPHY

[Altmann, 1980] Altmann, G. (1980). Prolegomena to Menzerath's law. Glottometrika 2, 1–10.

[Altmann, 2002] Altmann, G. (2002). Einführung in die quantitative Lexikologie. Göttingen: Peust & Gutschmidt.

[Altmann, 2012] Altmann, G. (2012). Certain differences between qualitative and quantitative linguistics. Czech and Slovak Linguistics Review 1, 6–15.

[Altmann, Schwibbe, Kaumanns, 1989] Altmann, G., Schwibbe, M.H., Kaumanns, W. (Eds.) (1989). Das Menzerathsche Gesetz in informationsverarbeitenden Systemen. Hildesheim: Olms.

[Andres, 2009] Andres, J. (2009). On de Saussure's principle of linearity and visualization of language structures. Glottotheory 2(2), 1–14.

[Andres, 2010] Andres, J. (2010). On a conjecture about the fractal structure of language. Journal of Quantitative Linguistics 17(2), 101–122.

[Andres, 2014] Andres, J. (2014). The Moran–Hutchinson formula in terms of Menzerath–Altmann's law and Zipf–Mandelbrot's law. In: Altmann, G., Čech, R., Mačutek, J., Uhlířová, L. (eds.): Empirical Approaches to Text and Language Analysis. Studies in Quantitative Linguistics 17, Lüdenscheid: RAM-Verlag, pp. 29–44.

[Andres, Benešová, 2011] Andres, J., Benešová, M. (2011). Fractal analysis of Poe's Raven. Glottometrics 21, 73–98.

[Andres, Benešová, 2012] Andres, J., Benešová, M. (2012). Fractal analysis of Poe's Raven, II. Journal of Quantitative Linguistics 19(4), 301–324.

[Andres et al., 2012a] Andres, J., Benešová, M., Kubáček, L., Vrbková, J. (2012). Methodological note on the fractal analysis of texts. Journal of Quantitative Linguistics 19(1), 1–31.

[Andres et al., 2012b] Andres, J., Kubáček, L., Machalová, J., Tučková, M. (2012). Optimization of parameters in the Menzerath–Altmann law. Acta Univ. Palacki. Olomuc., Fac. rer. nat., Math. 51(1), 5–27.

[Andres et al., 2014] Andres, J., Benešová, M., Chvosteková, M., Fišerová, E. (2014). Optimization of parameters in the Menzerath–Altmann law, II. Acta Univ. Palacki. Olomuc, Fac. rer. nat., Mathematica 53 (2), 5–28.

[Andres, Benešová, Langer, 2019] Andres, J., Benešová, M., Langer, J. (2019). Towards a fractal analysis of the sign language. J. Quantitative Linguist., 1–18 doi:10.1080/09296174.2019.1656149.

[Andres, Langer, Matlach, 2020] Andres, J., Langer, J., Matlach, V. (2020). Fractal-based analysis of sign language. Commun. Nonlin. Sci. Numer. Simul. 84, 105214, 1–14.

[Andres, Rypka, 2012] Andres, J., Rypka, M. (2012). Self-similar fractals with a given dimension and the application to quantitative linguistics. Nonlinear Analysis – B: Real World Applications 13(1), 42–53.

[Ausloos, 2012a] Ausloos, M., (2012). Generalized Hurst exponent and multifractal function of original and translate texts mapped into frequency and length time series. Phys. Rev. E 86, 1–9.

[Ausloos, 2012b] Ausloos, M. (2012). Measuring complexity with multifractals in texts. Translation effects. Chaos, Solitons & Fractals 45 (11), 1349–1357.

[Baayen et al., 2002] Baayen, R. H., Halteren, H. Van, Neijt, A., Tweedie F. (2002). An experiment in authorship attribution. In: Proceedings of JADT 2002, St. Malo, pp. 29–37.

[Baixeries et al., 2013] Baixeries, J., Hernández-Fernández, A., Forns, N., Ferrer-i-Cancho, R. (2013). The Parameters of Menzerath–Altmann Law in Genomes. Journal of Quantitative Linguistics 20(2), 94–104.

[Baláž et al., 1988] Baláž, J. et al. (1988). Interpersonálna komunikácia a jej poruchy [Interpersonal communication and its disorders]. Bratislava: SPN.

[Barnsley, 2013] Barnsley, M.F. (2013). Fractals Everywhere: New Edition. Mineola, N.Y: Courier Dover Publications.

[Battison, 1978] Battison, R. (1978). Lexical Borrowing in American Sign Language. Silver Springs: Linstok Press.

[Bellugi, Fischer, 1972] Bellugi, U., Fischer, S. (1972). A comparison of sign language and spoken language. Cognition 1(2-3), 173–200.

[Benešová, Faltýnek, Hadwiger, 2014] Benešová, M.; Faltýnek, D.; Hadwiger, L. (2014). Menzerath–Altmann law in differently segmented texts. In: Qualico 2014. Olomouc: Palacky University.

[Bianchi, Pianese, 2018] Bianchi, S., Pianese, A. (2018). Time–varying Hurst–Hölder exponents and the dynamics of (in)efficiency in stock markets. Chaos, Solitons and Fractals 109, 64–75.

[Biberauer, Roberts, 2013] Biberauer, T., Roberts, I. (Eds.) (2013). Challenges to Linearization. Berlin: De Gruyter.

[Bímová, 2002] Bímová, P. (2002). Jazyk znakový – jazyk přirozený [Sign Language – Natural Language]. Čeština doma a ve světě 10, 100–103.

[Binongo, 2003] Binongo, J.N.G. (2003). Who wrote the 15th book of Oz? An application of multivariate analysis to authorship attribution. Chance 16(2), 9–17.

[Bolshoy, 2003] Bolshoy, A. (2003). DNA sequence analysis linguistic tools: contrast vocabularies, compositional spectra and linguistic complexity. Applied bioinformatics 2, 103–112.

[Brennan et al., 1990] Brennan, M. et al. (1990). Words in hand. A structural analysis of the signs of British Sign Language. Edinburgh: Moray House College of Education.

[Brentari, 2011] Brentari, D. (2011). Handshape in Sign Language Phonology. In: Companion to Phonology, Oxford University Press, pp. 195–222.

[Brentari et al., 2011] Brentari, D., Coppola, M., Whan Cho, P., Senghas, A. (2017). Handshape Complexity as a Precursor to Phonology: Variation, Emergence, and Acquisition. Language Acquisition 24(4), 1–24.

[Burrows, 2003] Burrows, J. (2003). Questions of Authorships: Attribution and Beyond. Computersand the Humanities 37(1), 5–32.

[Chatterjee, Hadi, 2006] Chatterjee, S., Hadi, A. S. (2006). Regression Analysis by Example. Hoboken, New Jersey: John Wiley & Sons, Inc.

[Chomsky, 2002] Chomsky, N. (2002). Syntactic Structures. Walter de Gruyter.

[Colville, 1986] Colville, M.D. (1986). The Edinburgh non-manual coding system (ENCS). In: Tervoort, B.T. (ed.), Signs of life. Proceedings of the Second European Congress on Sign Language Research, Amsterdam July 14–18, 1985. Amsterdam: University of Amsterdam, pp. 204–208.

[Cramer, 2005] Cramer, I. (2005). The parameters of the Altmann–Menzerath law. Journal of Quantitative Linguistics 12(1), 41–52.

[Čermák, 2001] Čermák, F. (2001). Jazyk a jazykověda. Přehled a slovníky [Language and linguistics. Overview and dictionaries]. Praha: Karolinum.

[Černý, 1997] Černý, J. (1997). Úvod do studia jazyka [Introduction to the study of language]. Olomouc: Rubico.

[Červenková, 1997] Červenková, A. (1997) Výzkum znakového jazyka [Sign language research]. Info zpravodaj – magazín informačního centra o hluchotě 5(2), 6–8.

[Diederich, Kindermann, Paass, 2003] Diederich, J., Kindermann, E., Paass, G. (2003). Authorship attribution with support vector machines. Applied Intelligence, 19(1-2), 109–123.

[Damerau, 1964] Damerau, F.J. (1964). A technique for computer detection and correction of spelling errors. Communications of the ACM 7(3), 171–176.

[De Saussure, 2008] De Saussure, F. (2008). Kurs obecné lingvistiky. 3. vyd. [General linguistics course. 3rd edition]. Praha: Academia.

[Draves et al., 1964] Draves, S., Abraham, R., Viotti, P, Abraham, F.D., Sprott, J.C. (2008). The aesthetics and fractal dimension of electric sheep. Int. J. Bifurc. Chaos 18(4), 1243–1248.

[Drożdż et al., 2016] Drożdż, S., Oświęcimka, P., Kulig, A., Kwapień, J., Bazarnik, K., Grabska-Gradzińska, I., Rybicki, J., Stanuszek, M. (2016). Quantifying origin and character of long-range correlations in narrative texts. Information Sciences 331, 32–44.

[Eisenhauer, 2003] Eisenhauer, J.G. (2003). Regression through the Origin. Teaching Statistics 25, 76–80.

[Efron, Tibshirani, 1993] Efron, B., Tibshirani, R.J. (1993). An Introduction to the Bootstrap. Boca Raton, N.Y.: Chapman & Hall/CRC.

[Eftekhari, 2006] Eftekhari, A. (2006). Fractal geometry of texts: First attempt to Shakespeare's. Journal of Quantitative Linguistics, 13(2-3), 177–193.

[Engberg-Pedersen, 1986] Engberg-Pedersen, E. (1986). The Use of Space with Verbs in Danish Sign Language. In: Tervoort, B.T. (ed.), Signs of life. Proceedings of the Second European Congress on Sign Language Research, Amsterdam July 14–18, 1985. Amsterdam: University of Amsterdam, pp. 32–41.

[Engberg-Pedersen, Hansen, 1986] Engberg-Pedersen, E., Hansen, B. (1986). Workshop II, 1. 30 Signs from Danish Sign Language. In: Tervoort, B.T. (ed.), Signs of life. Proceedings of the Second European Congress on Sign Language Research, Amsterdam July 14–18, 1985. Amsterdam: University of Amsterdam, pp. 209–217.

[Eroglu, 2014] Eroglu, S. (2014). Self-organization of genic and intergenic sequence lengths in genomes: Statistical properties and linguistic coherence. Complexity 21(1), 268–282.

[Evans, 2001] Evans, L. (2001). Totální komunikace, struktura a strategie [Total communication: structure and strategy]. Hradec Králové: Pedagogické centrum.

[Feder, 1988] Feder, J. (1988). Fractals. New York: Plenum Press.

[Fedurek, Zuberbühler, Semple, 2017] Fedurek, P., Zuberbühler, K., Semple, S. (2017). Trade-offs in the production of animal vocal sequences: insights from the structure of wild chimpanzee pant hoots. Frontiers in Zoology 14, 50.

[Ferrer-i-Cancho, Elvevag, 2010] Ferrer-i-Cancho, R., Elvevag, B. (2010). Random Texts Do Not Exhibit the Real Zipf's Law–Like Rank Distribution. PLoS ONE, vol. 5, e9411.

[Ferrer-i-Cancho et al., 2013] Ferrer-i-Cancho, R., Forns, N., Hernández-Fernández, A., Bel-Enguix, G., Baixeries, J. (2013). The Challenges of Statistical Patterns of Language: The Case of Menzerath's Law in Genomes. Complexity 18(3), 11–17.

[Freeman, Carbin, Boese, 1992] Freeman, R.D., Carbin, C.F., Boese, R.J. (1992). Tvé dítě neslyší? [Can't Your Child Hear?] Praha: FRPSP.

[Frosini, 1991] Frosini, P. (1991). Measuring shapes by size functions. In: Proceedings of SPIE on Intelligent Robotics and Computer Vision X, Vol. 1607, Boston, pp. 3–26.

[Frosini, Landi, 2001] Frosini, P., Landi, C. (2001). Size functions and formal series. Applicable Algebra in Engin. Communic. Comput. 12, 327–349.

[Gregory et al., 2001] Gregory, S. et al. (2001). Problémy vzdělávání sluchově postižených [Issues in Deaf education]. Praha: FF UK.

[Gustison et al., 2016] Gustison, M.L., Semple, S., Ferrer-i-Cancho, R., Bergman, T.J. (2016). Gelada vocal sequences follow Menzerath's linguistic law. In: Dorothy L. Cheney (ed.): Proceedings of the National Academy of Sciences, vol. 113(19), E2750–E2758.

[Hadler, Irrgang, 2014] Hadler, F., Irrgang, D. (2014). Nonlinearity, multilinearity, simultaneity: notes on epistemological structures. In: (Moura, H., Sternberg, R., Cunha, R., Queiroz, C., Zeilinger, M. (eds.): Interactive Narratives, New Media & Social Engagement, Toronto: Univ. of Toronto, pp. 70–87.

[Handouyahia, Ziou, Wang, 1999] Handouyahia, M., Ziou, D., Wang, S. (1999). Sign language recognition using movement-based size functions. In: Proceedings of Vision Interface '99 (Trois-Rivières, Canada, 19–21 May), Canadian Image Processing and Pattern Recognition Society, pp. 210–216.

[Haspelmath, Tadmor, 2009] Haspelmath, M., Tadmor, U. (Eds.) (2009). Loanwords in the World's Languages: A Comparative Handbook. Berlin and New York: Mouton de Gruyter.

[Hattala, 1870] Hattala, M. (1870). Počátečné skupeniny souhlásek československých (Initial consonant clusters in Czechoslovak). Praha: Kr. české spol. nauk.

[Havlin et al., 1995] Havlin, S., Buldyrev, S.V., Goldberger, A.L., Mantegna, R.N., Peng, C.K., Simons, M., Stanley, H.E. (1995). Statistical and Linguistic Features of DNA Sequences. Fractals 3(2), 269–284.

[den Heijer, Eiben, 2010] Heijer, E., den, Eiben, A.E. (2010). Comparing aesthetic measures for evolutionary art. In: Di Chio, C. et al. (eds.): Applications of Evolutionary Computation. Proceedings of the Conference on Applications of Evolutionary Computation EvoApplications 2010: EvoCOMNET, EvoENVIRONMNET, EvoFIN, EvoMUSART, and EvoTRANSLOG (Istanbul, Turkey, April 2010), Part II, LNCS 6025, Berlin: Springer, pp. 311–320.

[Hockett, 1982] Hockett, C.F. (1982). The view from language: Selected Essays, 1948–1974. Language 58(3), 686–690.

[Hointing, Slobin, 2002] Hointing, N., Slobin, D.I. (2002). Transcription as a tool for understanding the Berkeley transcription system for sign language research (BTS) [online]. [2006-01-26]. Avaliable at: http://ihd.berkeley.edu/btsforsignlanguage.pdf.

[Holmes, Robertson, Paez, 2001] Holmes, D. I., Robertson, M., Paez, R. (2001). Stephen crane and the new-york tribune: A case study in traditional and non-traditional authorship attribution. Computers and the Humanities, 35(3), 315–331.

[Holubová, 2006] Holubová, M. (2006). Prostor v českém znakovém jazyce [Space in Czech sign language] [online]. [2012-07-15]. Avaliable at: http://ruce.cz/clanky/42-prostor-v-ceskem-znakovem-jazyce.

[Homoláč, 1998] Homoláč, J. (1998). Komunikace neslyšících. Sociolingvistika (antologie textů) [Communication of the deaf. Sociolinguistics (anthology of texts)]. Praha: FF UK.

[Hronová, 2002] Hronová, A. (2002). Poznáváme český znakový jazyk III. (Tvoření tázacích vět) [We get to know Czech sign language III. (Forming question sentences)]. Speciální pedagogika, 12(2), 113–123.

[Holubová, Motejzíková, 2002] Hronová, A.; Motejzíková, J. (2002). Raná komunikace mezi matkou a dítětem [Early communication between mother and child]. Praha: FRPSP.

[Hrubý, 1999] Hrubý, J. (1999). Průvodce neslyšících a nedoslýchavých po jejich vlastním osudu – 1. díl [Guide to the Deaf and Hard of Hearing of Their Own Destiny – Part 1]. 2nd edition, Praha: FRPSP.

[Hřebíček, 1990] Hřebíček, L. (1990). The constants of Menzerath–Altmann's law. Glottometrika 12, 61–71.

[Hřebíček, 1992] Hřebíček, L. (1992). Text in Communication: Supra-Sentence Structures. Bochum: Brockmeyer.

[Hřebíček, 1994] (1994). Fractals in language. Journal of Quantitative Linguistics 1(1), 82–86.

[Hřebíček, 1995] Hřebíček, L. (1995). Text Levels. Language Constructs, Constituents and the Menzerath–Altmann Law. Trier: Wissenschaftlicher Verlag.

[Hřebíček, 1997] Hřebíček, L. (1997). Lectures on Text Theory. Prague: The Academy of the Sciences of the Czech Republic (Oriental Institute).

[Hřebíček, 1998a] (1998). Hurst's indicators and text. Some properties of word–frequency series. In: Altmann, J., Koch, W. A. (eds.), Systems. New Paradigma for the Human Sciences. Berlin: W. de Gruyter, pp. 572–588.

[Hřebíček, 1998b] (1998). Language fractals and measurement in texts. Archiv orientální 66, 233–242.

[Hřebíček, 2000] Hřebíček, L. (2000). Variation in Sequences. Oriental Institute Prague, The Academy of Sciences of the Czech Republic, Prague.

[Hřebíček, 2002] Hřebíček, L. (2002). Vyprávění o lingvistických experimentech s textem [Stories about Linguistics Experiments with the Text]. Praha: Academia (in Czech).

[Hřebíček, 2007] Hřebíček, L. (2007). Text in Semantics. Prague: The Academy of Sciences of the Czech Republic (Oriental Institute).

[Hřebíček, 2007b] Hřebíček, L. (2007). Semantic slaps in text structures. Slovo a slovesnost 68, 83–90 (in Czech).

[Hřebíček, 2008] Hřebíček, L. (2008). Filologie versus lingvistika [Filology vs. linguistics]. Vesmír 87, 488, [online] [cit. 2011-08-23].

[Hurst, 1951] Hurst, H.E. (1951). Long–term storage capacity of reservoirs. Transactions 350 of American Society of Civil Engineers 116, 770–799.

[Hurst, Black, Simaika, 1965] Hurst, H.E., Black, R.F., Simaika, Y.M. (Eds.) (1965). Long–Term Storage: An Experimental Study. Constable, London.

[Hutchinson, 1981] Hutchinson, J.E. (1981). Fractals and self similarity. Indiana Univ. Math. J. 30, 713–747.

[Jabůrek, 2004] Jabůrek, J. (1998). Bilingvální vzdělávání neslyšících [Bilingual education of the deaf]. Praha: Septima.

[Janovcová, 2004] Janovcová, Z. (2004). Alternativní a augmentativní komunikace [Alternative and augmentative communication]. Brno: Masarykova univerzita.

[Jammer, 2006] Jammer, M. (2006). Concepts of Simultaneity: From Antiquity to Einstein and Beyond. Baltimore: The John Hopkins Univ. Press.

[Jašíčková, Benešová, Faltýnek, 2013] Jašíčková, A., Benešová, M., Faltýnek, D. (2013). An application of the Menzerath–Altmann law to a sample produced by an aphasic patient. Czech and Slovak Lingvistic Review 2.

[Johnston, Schembri, 2007] Johnston, T., Schembri, A. (2007). Australian Sign Language: An Introduction to Sign Language Linguistics. Cambridge: Cambridge University Press.

[Joshua et al., 2018] Joshua, J., Borneman, D., Malaia, E., Wilbur, R.B. (2018). Motion characterization using optical flow and fractal complexity. Journal of Electronic Imaging 27, 5, 051229, 1–6.

[Juola, 2006] Juola, P. (2006). Authorship attribution. Foundations and Trends in Information Retrieval 1(3), DOI: 10.1561/1500000005.

[Juola, Baayen, 2003] Juola, P., Baayen, H. (2003). A controlled-corpus experiment in authorship identification by cross-entropy. Literary and Linguistic Computing 20, 59–67.

[Käding, 1897] Käding, F.W. (1897). Häufigkeitswörterbuch der deutschen Sprache. Steglitz.

[Karlík, Nekula, Pleskalová et al., 2002] Karlík, P., Nekula, M., Pleskalová, J. et al. (Eds.) (2002). Encyklopedický slovník češtiny [Encyclopedic dictionary of Czech]. Praha: Lidové noviny.

[Karlík, Nekula, Pleskalová, 2020] Karlík, P., Nekula, M., Pleskalová, J. (Eds.) (2020). Nový encyklopedický slovník češtiny [New encyclopedic dictionary of Czech]. Praha: Lidové noviny.

[Klima, Bellugi, 1979] Klima, E.S., Bellugi, U. (1979). The Signs of Language. Cambridge: Harvard University Press.

[Köhler, 1982] Köhler, R. (1982). Das Menzerathsche Gesetz auf Satzebene. Glottometrika 4, 103–113.

[Köhler, 1984] Köhler, R. (1984). Zur Interpretation des Menzerathschen Gesetzes. Glottometrika 6, 177–183.

[Köhler, 1989] Köhler, R. (1989). Das Menzerathsche Gesetz als Resultat des Sprachverarbeitungsmechanismus. In: Altmann, G., Schwibbe, M.H. (eds.): Das Menzerathsche Gesetz in informations – verarbeitenden Systemen. Hildesheim: Olms, pp. 108–116.

[Köhler, 1995] Köhler, R. (1995). Maßeinheiten, Dimensionen und fraktale Strukturen in der Linguistik. Zeitschrift für Empirische Textforschung 2, 5–6.

[Köhler, 1997] Köhler, R. (1997). Are there fractal structures in language? Units of measurement and dimensions in linguistics. Journal of Quantitative Linguistics 4(1-3), 122–125.

[Köhler, 2008] Köhler, R. (2008). The fractal dimension in script: an experiment. In: Altmann, G., Feng Xiang, F. (eds.), Analysis of Script, Properties of Characters and Writing Systems. Berlin: de Gruyter, pp. 115–120.

[Köhler, 2014] Köhler, R. (2014). The fractal structure of linguistic motifs. In: Altmann, G. et al. (eds.), Empirical Approaches to Text and Language Analysis. Lidenscheid: RAM-Verlag, pp. 94–104.

[Köhler, Altmann, Piotrowski, 2005] Köhler, R., Altmann, G., Piotrowski, R.G. (Eds.), (2005). Quantitative Linguistics: An International Handbook. Berlin, New York: De Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft, 27).

[Köhler, Hoffmann, 1995] Köhler, R., Hoffmann, C. (Eds.) (1995). Bibliography of Quantitative Linguistics. Amsterdam: J. Benjamin Publ. Co.

[Koluchová, 1989] Koluchová, J. et al. (1989). Přehled patopsychologie dítěte II [Overview of child pathopsychology II]. Olomouc: Univerzita Palackého.

[Komenský, 1959] Komenský, J.A. (1959). Janua linguarum reserata. Praha: SPN.

[Krahulcová, 2002] Krahulcová, B. (2002). Komunikace sluchově postižených [Communication of the hearing impaired]. 2nd edition, Praha: Karolinum.

[Kremers, 2009] Kremers, J. (2009). Recursive linearization. The Linguistic Review 26(1), 135–166.

[Kremers, 2012] Kremers, J. (2012). The syntax of simultaneity. Lingua 122(9), 979–1003.

[Křivohlavý, 1988] Křivohlavý, J. (1988). Jak si navzájem lépe porozumíme [How we better understand each other]. Praha: Svoboda.

[Kubáček, Kubáčková, 2000] Kubáček, L., Kubáčková, L. (2000). Statistics and Metrology. Olomouc: Palacký University (in Czech).

[Kulacka, 2010] Kulacka, A. (2010). The coefficients in the formula for the Menzerath–Altmann law. Journal of Quantitative Linguistics 17(4), 257–268.

[Kulacka, Mačutek, 2007] Kulacka, A., Mačutek, J. (2007). A discrete formula for the Menzerath–Altmann law. Journal of Quantitative Linguistics 14(1), 23–32.

[Lane, Hoffmeister, Bahan, 1996] Lane, H., Hoffmeister, R., Bahan, B.A. (1996). Journey into the Deaf-World. San Diego: Dawn Sign Press.

[Langer, 2006] Langer, J. (2006). Výuka znakového jazyka a multimédia [Teaching sign language and multimedia]. In: Pedagogický software 2006. České Budějovice: Scientific Pedagogical Publishing, pp. 454–456. Avaliable at: http://www.uss.upol.cz/cz/clenove/profil/langer.

[Langer, 2007] Langer, J. (2007). Nová funkce slovníků znakového jazyka [A new feature of sign language dictionaries]. Gong 36(3), 15–16. Avaliable at: http://www.uss.upol.cz/cz/clenove/profil/langer.

[Langer, 2008] Langer, J. (2008). Obousměrný překlad v elektronických slovnících znakového jazyka [Bidirectional translation in electronic sign language dictionaries]. In: Pedagogický software 2008 [CD]. České Budějovice: Scientific Pedagogical Publishing, pp. 321–326. Avaliable at: http://www.uss.upol.cz/cz/clenove/profil/langer.

[Langer, 2013] Langer, J. (2013). Znakově systěmy v komunikaci neslyšících [Sign systems in communication of the Deaf]. Olomouc: Univerzita Palackého.

[Langer et al., 2020] Langer, J., Andres, J., Benešová, M., Faltýnek, D. (2020). Some aspects of a sign language quantitative analysis. To appear.

[Langer, Rypka, 2017] Langer, J., Rypka, M. (2017). Testing of subjective perception of complexity of signs of Czech sign language. Journal of Exceptional People 11(2), 123–141.

[Lawson, 1983] Lawson, L. (1983). Multichannel Signs. In: Kyle, J.G.; Woll, B. (eds.), Language in Sign: An International Perspective on Sign Language. London: Croom Helm, pp. 97–103.

[Leopold, 2001] Leopold, E. (2001). Fractal structure in language. The question of the imbedding space. In: Köhler, R., Uhlířová, I., Wimmer, G. (eds.), Text as a Linguistic Paradigm: Levels, Constituents, Constructs. Festschrift in honour of Luděk Hřebíček, Trier: Wissenschaftlicher Verlag, pp. 163–176.

[Lillo-Martin, Gajewski, 2014] Lillo-Martin, D., Gajewski, J. (2014). One grammar or two? Sign Languages and the Nature of Human Language. In: Wiley Interdisciplinary Reviews. Cognitive Science 5(4), 387–401.

[Lorenz, Andres, Franck, 2017] Lorenz, W.E., Andres, J., Franck, G. (2017). Fractal aesthetics in architecture. Applied Mathematics & Information Sciences 11(4), 971–981.

[Macurová, 1996] Macurová, A. (1996). Proč a jak zapisovat znaky českého znakového jazyka (Poznámky k diskusi) [Why and how to annotate signs of Czech Sign Language (Notes for discussion)]. Speciální pedagogika 6(1), 5–20.

[Macurová, 1998] Macurová, A. (1998). Nad slovenským sborníkem „Štylistika neverbálnej komunikácie" [Above the Slovak anthology "Stylistics of nonverbal communication"]. Speciální pedagogika 8(4), 56–59.

[Macurová, 2001] Macurová, A. (2001). Poznáváme český znakový jazyk. (Úvodní poznámky) [We get to know Czech sign language. (Introductory remarks)]. Speciální pedagogika 11(2), 69–75.

[Macurová, 2003] Macurová, A. (2003). Poznáváme český znakový jazyk IV. Vyjadřování času [We get to know Czech sign language IV. Expressing time]. Speciální pedagogika 13(2), 89–98.

[Macurová, 2008] Macurová, A. (2008). Dějiny výzkumu znakového jazyka u nás a v zahraničí [History of sign language research at home and abroad]. 2nd edition. Praha: Česká komora tlumočníků znakového jazyka.

[Macurová, Bímová, 2001] Macurová, A., Bímová, P. (2001). Poznáváme český znakový jazyk II. (Slovesa a jejich typy) [We get to know Czech sign language II. (Verbs and their types)]. Speciální pedagogika 11(5), 285–296.

[Macurová, Homoláčová, Ptáček, 1997] Macurová, A., Homoláčová, I., Ptáček, V. (1997). Výzkum komunikace neslyšících: český znakový jazyk [Deaf communication research: Czech sign language]. Speciální pedagogika 7(3), 1–15.

[Malaia, 2017] Malaia, E. (2017). Current and future methodologies for quantitative analysis of information transfer in sign language and gesture data. Behavioral and Brain Sciences 40, E63.

[Malaia, Borneman, Wilbur, 2016] Malaia, E., Borneman, J.D., Wilbur, R.B. (2016). Assessment of information content in visual signal: analysis of optical flow fractal complexity. Visual Cognition 24(3), 246–251.

[Malaia, Borneman, Wilbur, 2018] Malaia, E., Borneman, J.D., Wilbur, R.B. (2018). Information transfer capacity of articulators in American Sign Language. Language and Speech 61(1), 97–112.

[Mandelbrot, 1983] Mandelbrot, B.B. (1983). The Fractal Geometry of Nature. New York: Freeman and Comp.

[Mandelbrot, 1985] Mandelbrot, B.B., (1985). Self–affinity and fractal dimension. Physica Scripta 32, 257–260.

[Mandelbrot, 2000] Mandelbrot, B.B. (2000). Les objects fractals. Forme, hazard et dimension. Paris: Flammarion.

[Mandelbrot, 2007] Mandelbrot, B.B., (2007). The (Mis)Behavior of Markets, A Fractal View of 380 Risk, Ruin and Reward. New York: Basic Books.

[Mandelbrot, Wallis, 1968] Mandelbrot, B.B., Wallis, J.R., (1968). Noah, Joseph, and operational hydrology. Water Resour. Res. 4, 908–918.

[Manin, 2009] Manin, D. Yu. (2009). Mandelbrot's model for Zipf's law: Can Mandelbrot's model explain Zipf's law for language? Journal of Quantitative Linguistics, 16(3), 274–285.

[Mareš, Křivohlavý, 1995] Mareš, J., Křivohlavý, J. (1995). Komunikace ve škole [Communication at school]. Brno: Masarykova univerzita.

[Markov, 1913] Markov, A. A. (1913). Primer statisticheskogo issledovaniya nad tekstom "Evgeniya Onegina", illyustriruyuschijsvyaz' ispytanij v cep', Izvestiya Akademii Nauk, SPb, VI seriya, tom 7, 9 3, 153–162.

[Massone, Curiel, 2004] Massone, M.I., Curiel, M. (2004). Sign Order in Argentine Sign Language. Sign Language Studies 5(1), 63–93. Avaliable at: http://muse.jhu.edu.

[Mathesius, 1970] Mathesius, V. (1970). O potenciálnosti jevů jazykových [About potentiality of language phenomena]. In: Vachek, J. (ed.): U počátků pražské školy jazykovědné. Praha, p. 5–34.

[Matlach, Faltýnek, 2016] Matlach, V., Faltýnek, D. (2016). Báze nejsou písmena [Bases are not letters]. In: Studie z aplikované lingvistiky 7(1), 20–38.

[Meier, 2009] Meier, R.P. (2009) Why different, why the same? Explaining effects and non-effects of modality upon linguistic structure in sign and speech. In: Meier, R.P., Cormier, K., Quinto-Pozos, D., Modality and Structure in Signed and Spoken Languages. Cambridge: Cambridge University Press.

[Montemurro, 2001] Montemurro, M.A. (2001). Beyond the Zipf–Mandelbrot law in quantitative linguistics. Physica A, 300, 567–578.

[Montemurro, 2004] Montemurro, M.A. (2004). A generalization of the Zipf–Mandelbrot law in linguistics. In: Gell-Mann, M., Tsalis, C. (eds.), Nonextensive Entropy: Interdisciplinary Applications. Oxford: Oxford University Press, pp. 347–356.

[Montemurro, Pury, 2002] Montemurro, M.A., Pury, P.A. (2002). Long-range fractal correlations in literary corpora. Fractals 10(4), 451–461.

[Montgomery, Peck, Vinig, 2006] Montgomery, D.C., Peck, E.A., Vinig, G.G. (2006). Introduction to Linear Regression Analysis. Fourth Edition, Wiley Series in Probability and Statistics. Hoboken, New Jersey: John Wiley & Sons, Inc.

[Moran, 1946] Moran, P.A.P. (1946). Additive functions of intervals and Hausdorff measure. Proc. Cambridge Philos. Soc. 42, 15–23.

[Mosteller, Wallace, 1964] Mosteller, F., Wallace, D. L. (1964). Inference and disputed authorship The Federalist. Reading, Mass: Addison-Wesley.

[Motejzíková, 2007] Motejzíková, J. (2007) Simultánnost v českém znakovém jazyce [Simultaneity in Czech sign language]. Diploma thesis. Praha: Univerzita Karlova.

[Mrzílková, 1996] Mrzílková, M. (1996) Neslyšící – členové kultury a společenství neslyšících [Deaf—members of the culture and community of the deaf]. Speciální pedagogika 6(1), 21–24.

[Natsopoulos, Abadzi, 1986] Natsopoulos, D., Abadzi, H. (1986). Understanding linguistic time sequence and simultaneity: A literature review and some new data. J. Psycholinguistic Research 15(3), 243–273.

[Němec, 1971] Němec, I. (1971). Hláskoslovné předpoklady produktivity slovotvorných formantů v češtině (The Phonological Prerequisites for the Productiveness of Word-Formative Elements in Czech). Slovo a slovesnost 32(3), 217–224.

[Nikolaou, 2014] Nikolaou, C. (2014). Menzerath–Altmann law in mammalian exons reflects the dynamics of gene structure evolution. Comput Biol Chem, Dec, 53, A, p. 134–143.

[Niyogi, Berwick, 1995] Niyogi, P., Berwick, R. C. (1995). A note on Zipf's Law, Natural Languages, and Noncoding DNA Regions. A. I. Memo, No. 1530, C.B.C.L. Paper No. 118.

[Nováková, Šebková, 2008] Nováková, R., Šebková, H. (2008). Číslovky v českém znakovém jazyce – praktická cvičení [Numerals in Czech sign language—practical exercises]. Praha: Česká komora tlumočníků znakového jazyka.

[Okrouhlíková, Šebková, 2008] Okrouhlíková, L. (2008). Lexikografie a dostupné slovníky znakované češtiny [Lexicography and available dictionaries of signed Czech]. 2nd edition, Praha: vCeská komora tlumočníků znakového jazyka.

[Okrouhlíková, Šebková, 2012] Okrouhlíková, L. (2012). Notace – zápis českého znakového jazyka [Notation of Czech sign language]. Dissertation thesis, Pedagogická fakulta, Praha: Univerzita Karlova.

[Pázman, 1988] Pázman, A. (1988). Foundations of Optimum Experimental Design. Dordrecht–Boston–Lancaster–Tokyo: D. Reidel Publishing Comp.

[Peitgen, Jürgens, Saupe, 2004] Peitgen, H.O., Jürgens, H., Saupe, D., (2004). Chaos and Fractals. New Frontiers of Science. 2nd ed., Berlin: Springer.

[Piantadosi, 2014] Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. Psychonomic Bulletin & Review 21, 1112–1130.

[Prün, 1994] Prün, C. (1994). Validity of Menzerath–Altmann's law: graphic representation of language, information processing systems and synergetic linguistics. Journal of Quantitative Linguistics 1(2), 148–155.

[Rudman, 1998] Rudman, J. (1998). The state of authorship attribution studies: Some problems and solutions. Computers and the Humanities 31, 351–365.

[Šebková, 2005] Šebková, H. (2005). Číslovky v českém znakovém jazyce [Numerals in Czech sign language]. Bachelor thesis, Praha: Univerzita Karlova.

[Servusová, 2008] Servusová, J. (2008). Kontrastivní lingvistika – český jazyk × český znakový jazyk [Contrastive linguistics—Czech language × Czech sign language]. 2nd edition. Praha: vCeská komora tlumočníků znakového jazyka.

[Sevinç, 2015] Sevinç, A.M. (2015). Finite-State Sign Language Morphophonology. PhD thesis, Graduate School of Informatics, Middle East Technical University, Ankara.

[Seydler, 1886] Seydler, A. (1886). Počet pravděpodobnosti v přítomném sporu [Probability calculus in a present conflict]. Athenaeum 3(7); 299–307.

[Sgall et al., 1973] Sgall, P., Bémová, A., Benešová, E., Goralčíková, A., Hajičová, E., Machová, S., Panevová, J., Piťha, P., Říha, A., Vomáčka, I., Weisheitelová, J. (1973). Úvod do algebraické lingvistiky [Introduction to Algebraic Linguistics]. Praha: SPN.

[Shahzad, Mittenthal, Caetano-Anollés, 2015] Shahzad, K., Mittenthal, J. E., Caetano-Anollés, G. (2015). The organization of domains in proteins obeys Menzerath–Altmann's law of language. BMC Systems Biology 9(44), 1–13.

[Slánská-Bímová, Okrouhlíková, 2008] Slánská-Bímová, P., Okrouhlíková, L. (2008) Rysy přirozených jazyků – Český znakový jazyk jako přirozený jazyk, Lexikografie – Slovníky českého znakového jazyka [Features of natural languages—Czech sign language as a natural language, Lexicography—Dictionaries of Czech sign language]. 2nd edition, Praha: Česká komora tlumočníků znakového jazyka.

[Slobin et al., 2001a] Slobin, D. et al. (2001). Sign language transcription at the level of meaning components: The Berkeley Transcription System (BTS). Sign language and linguistics 4(1–2), 63–104.

[Slobin et al., 2001b] Slobin, D. et al. (2001). The Berkeley Transcription System (BTS) Manual [online]. [cit. 2006-01-26]. Avaliable at: http://childes.psy.cmu.edu/manuals/bts.pdf.

[Souralová, 2004] Souralová, E. (2004). Znakový jazyk v komunikaci hluchoslepých [Sign language in deafblind communication]. Habilitation thesis, Olomouc: Univerzita Palackého.

[Štekauer, 2000] Štekauer, P. et al. (2000). Rudiments of English Linguistics. Prešov: Slovacontact.

[Stewart, 2014] Stewart, J. (2014). A quantitative analysis of sign lengthening in American Sign Language. Sign Language & Linguistics 17(1), 82–101.

[Stokoe, 1960] Stokoe, W.C. (1960). Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. Studies in linguistics: Occasional papers (No. 8).

[Stokoe, 2005] Stokoe, W.C., (2005). Sign language structure: an outline of the visual communication systems of the American deaf. The Journal of Deaf Studies and Deaf Education 10, 3–37.

[Strnadová, 2008] Strnadová, V. (2008). Specifické neverbální projevy neslyšících lidí [Specific non-verbal expressions of deaf people]. 2nd edition, Praha: Česká komora tlumočníků znakového jazyka.

[Svoboda, Krejčířová, Vágnerová, 2001] Svoboda, M., Krejčířová, D., Vágnerová, M. (2001). Psychodiagnostika dětí a dospívajících [Psychodiagnostics of children and adolescents]. Praha: Portál.

[Swadesh, 1955] Swadesh, M. (1955). Towards greater accuracy in lexicostatistic dating. International Journal of American Linguistics 21, 121–137.

[Tarcsiová, 2005] Tarcsiová, D. (2005). Komunikačný systém sluchovo postihnutých a spôsoby prekonávania ich komunikačnej bariéry [Communication system for the hearing impaired and ways to overcome their communication barrier]. Bratislava: Sapientia.

[Tarcsiová, 2008] Tarcsiová, D. (2008). Pedagogika sluchovo postihnutých [Pedagogy of the hearing impaired]. Bratislava: MABAG.

[Tetauerová 2008] Tetauerová, I. (2008). Mezinárodní znakový systém [International sign system]. 2nd editioin, Praha: Česká komora tlumočníků znakového jazyka.

[Těšitelová, 1987] Těšitelová, M. (1987). Kvantitativní lingvistika [Quantitative Linguistics]. Praha: SPN (in Czech).

[Thode, 2002] Thode, H.C. (2002). Testing for Normality. New York: Marcel Dekker.

[Tsonis, Elsner, Tsonis, 1997] Tsonis, A.A., Elsner, J.B., Tsonis, P.A. (1997). Is DNA a Language? Journal of Theoretical Biology 184(1), 25–29.

[Tučková, 2010] Tučková, M. (2010). Optimal Design of Regression Experiments. Diploma Thesis, Olomouc: Palacký University (in Czech).

[Tuldava, 1995] Tuldava, J. (1995). Methods in Quantitative Linguistics. Trier: Wissenschaftlicher Verlag.

[Universität Hamburg, 2020] Universität Hamburg (2020). Sign Language Notation System [online]. [2012-08-02]. Avaliable at: https://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/welcome.html.

[Uras, Verri, 1994] Uras, C., Verri, A. (1994). On the recognition of the alphabet of the sign language through size functions. In: Proceedings 12th IAPR International Conference on Pattern Recognition, Jerusalem, Vol. II, IEEE Comp. Soc. Press, Los Alamitos, pp. 334–338.

[Uras, Verri, 1995] Uras, C., Verri, A. (1995). Sign language recognition: an application of the theory of size functions. In: Proceedings of the 6th British conference on Machine vision (Vol. 2), BMVA Press Surrey, Birmingham, UK, pp. 711–720.

[Vašák, 1980] Vašák, P. (1980): Metody určování autorství literárních děl [Methods of Author Assigning of Literary Works]. Praha: Academia.

[Vermeerbergen, Leeson, Crasborn, 2007] Vermeerbergen, M., Leeson, L., Crasborn, O.A. (Eds.) (2007). Simultaneity in Signed Languages: Form and Function. Volume 28 of Current Issues in Linguistic Theory. Amsterdam: John Benjamins Publishing Company.

[Volek, 2002] Volek, J. (2002). Úvod do komunikačních studií [Introduction to communication studies]. Brno: Masarykova univerzita.

[Vybíral, 2005] Vybíral, Z. (2005). Psychologie komunikace [Psychology of communication]. Praha: Portál.

[Vysuček, 2008] Vysuček, P. (2008). Specifické znaky v českém znakovém jazyce [Specific signs in Czech sign language]. Praha: Česká komora tlumočníků znakového jazyka, (soubor).

[Weron, 2002] Weron, R., (2002). Estimating long–range dependence: finite sample prop erties and confidence intervals. Physica A: Statistical Mechanics and its Applications 312, 285–299.

[White, 1980] White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48(4), 817–838.

[Wimmer, Altmann, 2005] Wimmer, G., Altmann, G. (2005). Unified derivation of some linguistic laws. In: Köhler, R., Altmann, G., Piotrowski, R.G. (eds.), Quantitative Linquistics. An International Handbook. Berlin: De Gruyter, pp. 791–807.

[Wimmer et al., 2003] Wimmer, G., Altmann, G., Hřebíček, L., Ondrejovič, S., Wimmerová, S. (2003). Introduction to the Analysis of Texts. Bratislava: Veda (in Slovak).

[Yule, 1938] Yule, G. U. (1938). On sentence-length as a statistical characteristic of style in prose, with application to two cases of disputed authorship. Biometrika 30, 363–90.

[Yule, 1944] Yule, G. U. (1944). The Statistical Study of Literary Vocabulary. Cambridge: Cambridge University Press.

[Zhao, Zobel, 2005] Zhao, Y., Zobel, J. (2005). Effective and scalable authorship attribution using function words. In: Proceedings of 2nd Asian Information Retrieval Symposium, 3689 of LNCS, pp. 174–189.

[Zipf, 1949] Zipf, G. K. (1949). Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology. Addison-Wesley Press.

Peter Angermann, preparatory drawing for animation of *A. I. (Nohandsfractal)*, 2001

AUTHORS' PROFILES

Doc. Mgr. Jiří Langer, Ph.D. (born 1979 in Šternberk) is an associate professor at the Institute of Special Education Studies, Faculty of Education, Palacký University in Olomouc. He has been the deputy director of this institute since 2009. In addition to this, since 2012 he has also held the position of deputy director of the Palacký University Support Centre for Students with Special Needs. Jiří Langer is also active in academic policy; he has been a member of the Academic Senate of Palacký University in Olomouc since 2002 and since 2014 he has acted as chairman of this Senate. He has also been invited to give lectures at several European and Chinese universities and helped to establish a support centre for students with special needs at Leshan Normal University in China. In his research, Jiří Langer is interested in special education, especially in communication for Deaf and Deafblind people. He mainly focuses on Czech sign language, its education, lexicography, and interpreting. His colleague Eva Souralová and he are pioneers in the research of Czech tactile sign language for people with Deafblindness and have made considerable contributions towards its expansion in the Czech Republic. Jiří Langer is the main author of eleven electronic dictionaries of Czech sign language, and in 2006 he published probably the first electronic bidirectional dictionary between spoken and sign language *Slovník pojmů znakového jazyka pro oblast tělesné výchovy a sportu [Sign Language Dictionary – Sport and Physical Education]* (Fortuna, Praha, 2006). He has also published dozens of study texts, research papers, he has also authored dictionary sections of textbooks for pupils with hearing impairment, and is a co-author of the text *Act on the Communication Systems of Deaf and Deafblind People* (2008).

Prof. RNDr. dr hab. Jan Andres, DSc. (born 1954 in Olomouc) serves as a full professor at the Department of Mathematical Analysis and Applications of Mathematics, Faculty of Science, Palacký University in Olomouc. For the period of 1994–2003, he was a vice-dean at the Faculty of Science and from 2003–2018 he was the head of the Department of Mathematical Analysis and Applications of Mathematics. He has also acted as a visiting professor at two American universities and many European universities, among those the University of Rome 1 – La Sapienza and the University of Paris 1 – Sorbonne. Jan Andres sits on the board of numerous international scientific journals. In his research, he mainly focuses on nonlinear and multivalued analysis, differential equations and inclusions, dynamical systems and fractals, and their application to quantitative linguistics. He is the co-author (with Lech Górniewicz) of *Topological Fixed Point Principles for Boundary Value Problems*, a monograph published in 2003 by Kluwer; the second edition was published in 2012 by Springer, jointly with World Publishing Corporation, in Beijing. He is also the author of two invited chapters in handbooks, published in 2005 by Springer and in 2006 by Elsevier, and more than 200 scientific papers. Jan Andres writes poetry, philosophical essays, and articles on fine art.

Mgr. Martina Benešová, Ph.D. (born 1975 in Jihlava) has been a member of the Department of General Linguistics, Faculty of Arts, Palacký University in Olomouc since its foundation. She is interested in the overlaps of linguistics and arts with mathematics, which allows her to understand and grasp the beauty of a language and arts in a deeper way. Using the tools of quantitative linguistics, she has been studying translations of the poem *Raven* by Edgar Allan Poe into various languages. In her opinion, attempts like this make linguistic laws more precise, in an exact scientific manner. She is also a co-editor of *Recent Contributions to Quantitative Linguistics* (De Gruyter Mouton, Berlin, 2015) and a co-author of the monograph titled *Text Segmentation for Menzerath–Altmann Law Testing* (Palacký University, Olomouc, 2016). Her research has been mainly stimulated by the ideas of Luděk Hřebíček. Martina Benešová is also very fond of geology and mineralogy.

Mgr. Dan Faltýnek, Ph.D. (born 1980 in Olomouc) is an assistant professor at the Department of General Linguistics, Faculty of Arts, Palacký University in Olomouc. He served as the head of this department for the period of 2012–2016. Dan Faltýnek is interested in general linguistics, biosemiotics, and quantitative linguistics. His monograph *Co je nového v lingvistice [What's new in linguistics]* (Nová beseda, Praha, 2017) informs readers about the current state of linguistics and its interdisciplinary overlaps with the studies of structures represented by texts. His joint book with Vladimír Matlach, titled Gramatiky DNA [DNA Grammars] (Palacký University, Olomouc, 2014), reflects on linguistic methods for DNA analysis. His book, *Součet s nulou [Count Zero]* (Palacký University, Olomouc, 2014), illustrated by Klára Jastřembská, concerns the typology of zero categories like morphemes, elisions, ellipses, etc. One of the main topics of Faltýnek's research is the description of the protosemiosis, especially the structure of genetic code, which seems to be, according to his opinion, analogous to those one in sign languages, which however has not yet been clarified. Dan Faltýnek also writes poetry.
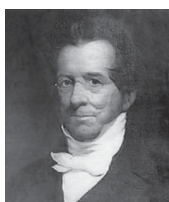
KEY FIGURES

Gabriel Altmann (1931–2020)
A German linguist, born in Slovakia, who was a main founder of quantitative linguistics.

Abbé Charles-Michel de l'Épée (1712–1789)
A French priest and educator, known as "Father of the Deaf", was a co-founder of the educational system of the deaf.

Thomas Hopkins Gallaudet (1787–1851)
An American educator, co-founder of the first permanent school for the deaf in North America. Gallaudet University was named in honor of him in 1894.

Luděk Hřebíček (1934–2015)
A Czech turkologist and figure one of the Czech quantitative linguistics.

Harold Edwin Hurst (1880–1978)
A British hydrologist who studied the long-term storage capacity of water reservoirs. His ideas were extended to other fields like finance, cardiology and quantitative linguistics.
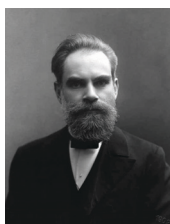
Aleksandr Mikhailovich Lyapunov (1857–1918)
A Russian mathematician, mechanician and physicist, who was a main founder of stability theory. His ideas have been also used in the theory of deterministic chaos.

Alena Macurová (1946)
A Czech bohemist and figure one in contemporary Czech sign language studies.

Benoit B. Mandelbrot (1924–2010)
A French–American mathematician, born in Poland, who was a main founder of fractal geometry. Its applications concern also quantitative linguistics.

Paul Menzerath (1883–1954)
A German phonetician who had observed for the first time that longer words tend to contain shorter syllables.

Ferdinand de Saussure (1857–1913)
A Swiss linguist and semiotician, regarded as a precursor of structuralism and quantitative linguistics.

William C. Stokoe, Jr. (1919–2000)
An American linguist whose research was pivotal in proving sign language as a true language. His study "Sign Language Structure" (1960) was the first linguistic analysis of the (American) sign language.

George Kingsley Zipf (1902–1950)
An American linguist and philologist who studied statistical occurrences in languages.

SPECIAL ACKNOWLEDGEMENTS

Our book is illustrated mainly by the series of pictures titled *Artificial Intelligence (Nohandsfractal)* by Peter Angermann (1945). This series started in 1989 and has continued until the present time.

For Angermann, the idea of artificial intelligence can be understood in the same way as the one of homunculus with a self-reproducing ability. Its more advanced version, a robot or a network, might be able to reproduce itself, forming a strange loop in the sense of Douglas R. Hofstadter. This recursion can be freely interpreted as a process with the self-similarity property, i.e. a fractal.

In the context of sign language, Angermann's illustrations are related both to our fractal analysis of sign language texts and manual aspects of a sign language itself.

The authors are grateful to Professor Angermann for a kind permission to reproduce his pictures.

Peter Angermann

Peter Angermann (born 1945, Rehau in Bavaria) is a neoexpressionist painter. During his studies from 1966 to 1968 at the Kunstakademie Nuremberg with Gerhard Wendland, and from 1968 to 1972 at the Kunstakademie Düsseldorf with Joseph Beuys, he was a co-founder of the group YIUP and together with Milan Kunc and Jan Knap of the group NORMAL. After a stay with his family in the USA in 1982 and 1983, in spring 1986 he taught as a guest lecturer for painting at an art school in Reykjavik, then from 1992 to 1993 at the Gesamthochschule Kassel. From 1996 to 2002 he was a Professor of painting at the Städelschule in Frankfurt and from 2002 to 2010 a Professor of painting at the Akademie der Bildenden Künste in Nuremberg. Besides his artistic work, he has also been intensely occupied with scientific questions, above all perception-theoretical and epistemological ones.

His most important one-man exhibitions were in 2013: *Light on the Horizon*, in the Museum Haus Lange in Krefeld; and in 2014: *The Desire for Seeing*, in the Neues Museum in Nuremberg. Important group exhibition participation includes 1980: *Times Square Show* in NYC, the *11th Biennale de Paris*, Musée d'Art Moderne; in 1984: *von hier aus* in Düsseldorf and *Tiefe Blicke* in Hessisches Landesmuseum in Darmstadt; in 2005: *Group Normal in the Prague Bien-*

nal; in 2014: *Wild Heart: German Neo-Expressionism since the 1960s*, China Art Museum, Shanghai; and in 2015: *DIE 80ER*, in Städel Museum in Frankfurt. In 2016, Professor Angermann was awarded by the prize of the city of Nuremberg.

Peter Angermann with his painting *Artificial Intelligence (Nohandsfractal)* from 2010

Jiří Langer
Jan Andres
Martina Benešová
Dan Faltýnek

Quantitative linguistic analysis
of Czech sign language

The main aim of this book is to present current research outcomes from quantitative analysis of Czech sign language. A multidisciplinary research project entitled "The Theoretical Basis for Teaching Czech Sign Language Tested through Quantitative Linguistic Methods" was carried out by researchers from three faculties of Palacký University in Olomouc, Czech Republic. It is the first attempt in the field of quantitative linguistics applied to a sign language. The authors believe that their book can serve at least as an introduction for further steps in this meritorious interdisciplinary area.

Palacký University
Olomouc