

Petr Sgall

Language in its multifarious aspects

CHARLES UNIVERSITY IN PRAGUE
THE KAROLINUM PRESS

Language in its multifarious aspects

Petr Sgall

Reviewed by

František Čermák, Ludmila Uhlířová

Edited by Eva Hajičová and Jarmila Panevová

Cover by Jan Šerých

Layout by Kamila Schüllerová

Typeset by DTP Nakladatelství Karolinum

First Edition

© Charles University in Prague, 2006

© Petr Sgall, 2006

ISBN 80-246-1158-9

ISBN 978-80-246-2547-8 (online : pdf)



Charles University in Prague
The Karolinum Press 2014

<http://www.cupress.cuni.cz>

LIST OF CONTENTS

Introduction by Eva Hajičová and Jarmila Panevová	7
A. General and Theoretical Issues	21
1. Types of languages and the simple pattern of the core of language (2003)	22
2. Freedom of language Its nature, its sources, and its consequences (2002)	44
3. On comparison of approaches (Remarks and illustrations) (2000)	67
4. Functionalism in Czech linguistics and in the world (1997)	80
5. Structure, meaning and use (1997)	95
6. Formal and Computational Linguistics in Prague (1995)	118
7. Underlying structure of sentences and its relations to semantics (1992)	130
8. A dependency based specification of topic and focus II – Formal account (1980)	140
9. Generative Beschreibung und die Ebenen des Sprachsystems (1964)	164
10. Introduction to Linguistic Morphology (1989)	182
B. Syntax	189
11. Underlying Structures in Annotating Czech National Corpus (2001)	190

12. Revisiting the classification of the dependents (1998)	200
13. Case and meaning (1980)	212
C. Topic-focus articulation	227
14. From functional sentence perspective to topic-focus articulation (2003)	228
15. The position of Czech linguistics in theme-focus research (1987)	238
16. Wortfolge und Fokus im Deutschen (1982)	248
17. Functional sentence perspective in a generative description (1967)	275
D. From sentence to discourse in semantics	303
18. Dynamics in the meaning of sentence and of discourse (2003)	304
19. From meaning via reference to content (1995)	325
20. Meaning, reference and discourse patterns (1994)	335
E. Typology of languages	363
21. Natürlichkeit, Syntax und Typologie (1998)	364
22. Die Sprachtypologie V. Skaličkas (1979)	375
23. On the notion "type of language" (1971)	394
24. Zur Typologie des Infinitivs (1958)	410
F. Speaking and writing	419
25. Spoken Czech revisited (2002)	420
26. Towards a theory of phonemic orthography (1987)	430
References	458
Bibliography of Petr Sgall	492
Name Index	544
Subject Index	551

INTRODUCTION

Petr Sgall (born May 27th, 1926 in České Budějovice, but spending most of his childhood in the small town Ústí nad Orlicí in eastern Bohemia and living since his university studies in Prague) is one of the most prominent Czech linguists belonging to the so-called “second generation” of the world-famous structural and functional Prague School of Linguistics. His first research interests focused on typology of languages, in which he was a pupil of Vladimír Skalička. His PhD thesis was on the development of inflection in Indo-European languages (published in Czech in 1958). He spent a year of postgraduate studies in Cracow, studying with J. Kuryłowicz. He habilitated as docent (associate professor) of general and Indoeuropean linguistics at Charles University in 1958 on the basis of his Cracow study of infinitive in Old Indian (*Infinitive im Rgveda*, published the same year). Since his beginnings, he was always deeply interested in the exceptional situation of Czech where alongside with the standard form of language there exists a form of Czech that is usually called ‘Common Czech’ (as it is not restricted to some geographical area as dialects are) and that is used by most Czech speakers in everyday communication. In this he was influenced by the work of Bohuslav Havránek on functional stratification of Czech.

At the beginning of the 1960s, Sgall was one of the first European scholars who got acquainted with the emerging new linguistic paradigm, Chomskyan generative grammar. On the one hand, he immediately understood the importance of an explicit description of language, but at

the same time, he was aware that the generative approach as presented in the early days of transformational grammar, lacks a due regard to the functions of language (at this point we want to recall his perspicacious analysis of Prague School functionalism in his paper published in 1964 in the renewed series Prague Linguistic Circle Papers (pre-war TLCP), the *Travaux linguistiques de Prague* Vol. I in 1964. Based on the Praguian tenets, Sgall formulated and developed an original framework of generative description of language, the so-called Functional Generative Description (FGD). His papers in the early sixties and his book presenting FGD (Sgall 1967) were the foundation stones of an original school of theoretical and computational linguistics that has been alive and flourishing in Prague since then. Sgall's innovative approach builds on three main pillars: (i) dependency syntax, (ii) information structure as an integral part of the underlying linguistic structure, and (iii) due regard to the distinction between linguistic meaning and cognitive content.

Petr Sgall has proved also outstanding organizational skills. In 1959, he founded a small subdepartment of mathematical linguistics (called then 'algebraic', to get distinguished from the traditional quantitative linguistics) and theory of machine translation at the Faculty of Arts of Charles University, followed by the foundation of a small group of computational linguistics also at the Faculty of Mathematics and Physics (in 1960) of the same University. In 1968, the two groups were integrated under his leadership into the Laboratory of Algebraic Linguistics, attached to the Faculty of Arts. This Laboratory, due to the political changes in the country caused by Russia-led invasion, had, unfortunately, a very short life-span. In 1972, Sgall faced a forced dismissal from the University for political reasons, and the whole group was eventually doomed to be dissolved. Fortunately, thanks to a group of brave colleagues and friends at the Faculty of Mathematics and Physics, he and his collaborators were transferred to this Faculty, less closely watched (by guardians of ideology) than was the domain of the Humanities. Even there, however, the conditions were not at all easy for him – for several years, the Communist Party decision for the group to disappear was in power, the number of Sgall's collaborators was harshly reduced and many obstacles were laid in the way of research in computational linguistics as such. Sgall himself was deprived of possibilities to teach, supervise students, travel to the West,

attend conferences there, and only slowly and gradually he could resume some of his activities in the 1980s. Nevertheless, not only the core of the research group continued working in contact with Western centers and their leading personalities (as evidenced above all by the contributions to his *Festschrift* edited by Jacob Mey and published by John Benjamins in 1986), but it was also possible to help three other immediately endangered colleagues to survive at the University.

The years after the political changes in our country in 1989 have brought him a due satisfaction after the previous years of suppression: a possibility of a 5-month stay as a research fellow at the Netherlands Institute of Advanced Studies in Wassenaar (a standing invitation he has had for many years but which he was not allowed to accept for political reasons), the membership in the prestigious *Academia Europaea*, the International Research Prize of Alexander von Humboldt in 1992, a visiting professorship at the University in Vienna in 1993, the Prize of the Czech Minister of Education in the same year, a honorary doctorate at the *Institut National des Langues et Civilisations Orientales* in Paris in 1995 and at the Hamburg University in 1998 and an honorary membership in the Linguistic Society of America in 2002, not to speak about numbers of invitations for lectures and conferences in the whole world, from the U.S.A. to Malaysia and Japan. As a Professor Emeritus of Charles University since 1995, he is still actively involved in teaching and supervising PhD students, in participating at Czech and international research projects and in chairing the Scientific Board of the Vilém Mathesius Center he helped to found in 1992.

Petr Sgall was also among those who helped to revive the Prague Linguistic Circle already in 1988 and has a substantial share in reviving also the book series *Travaux de Cercle linguistique de Prague* (under a parallel title *Prague Linguistic Circle Papers*), the first volume of which appeared in 1995 (published in Amsterdam by John Benjamins Publ. Company) and the fifth volume is now in preparation.

With his research activities based on a true Praguian functional approach, he thus more than made up for his negative attitudes published in the beginning of the fifties, a revolutionary and rash approach to which he was inspired by his wartime experience (his father died in Auschwitz, as did eleven of his closest relatives, and Petr Sgall himself spent some

months in a labour camp) and ill-advised by some of his tutors. Let us remind in this connection e.g. his review of three American volumes devoted to the Prague School published in 1978 in the Prague Bulletin of Mathematical Linguistics (a University periodical founded by Sgall in 1964), at the time when the political situation in the country and his own personal position was very difficult.

The present volume is conceived of as a reflection of the broad scope of Petr Sgall's linguistic interests, and, at the same time, as a document how lively the Prague School tenets are if developed by such a creative personality. Also, the contributions included in the volume illustrate characteristic features of Petr Sgall as a researcher: the overwhelming variety of deeply rooted topics of interest, the ability to penetrate into the substance of arguments and giving a convincing counterargument, the consistence of opinions but, at the same time, openmindedness and openness to discussion and willingness to accept the opponent's viewpoint if he finds good reasons for it. There are not many researchers of his position who would be able to react so creatively to stimuli from the outside, to learn a lesson from them and to push his students to do the same ('read if you want to be read' is one of his favourite slogans).

Sgall's papers selected for this volume have been sorted in six parts covering both general theoretical questions of language typology, linguistic description, relationships of grammar, meaning and discourse as well as more specific topics of the sentence structure and semantics. It is a matter of course that we could not omit at least a small sample of contributions to his most beloved child, functional stratification of Czech and orthography. Below, we give a very brief outline of the main views as present in the papers; we refer to the individual papers by their serial numbers in brackets.

Part A (**General and Theoretical Issues**) provides a broader picture of Sgall's understanding of the tenets of Prague School Linguistics and their reflection in the present-day development of language theories, including a brief characterization of the Functional Generative Description, based on a perspicuous account of the topic-focus articulation and on dependency syntax (4). Sgall has always been aware of the usefulness of comparison of linguistic frameworks and approaches (3). His original formal approach called Functional Generative Description (FGD) was

presented in a comparative perspective in the context of M. A. K. Halliday's Systemic (Functional) Grammar (5). FGD was proposed as early as in the mid-sixties (9) and was conceived of as an alternative to Chomskian generative transformational grammar. It is based on the dependency approach to syntax (8; this paper, in spite of its title, presents a proposal how to generate underlying dependency structures and is not concerned only with topic-focus articulation) and on a firm conviction that what constitutes the syntax of the sentence is its underlying structure rather than its surface shape (7). As a founder of computational linguistics in Prague (and in the whole of former Czechoslovakia), he has always been very sensitive to put a right balance to the formal and empirical aspects of that interdisciplinary domain (6). In this connection it should be recalled that Petr Sgall used his involuntary shift from the Faculty of Arts to the Faculty of Mathematics and Physics in the years after the Russian invasion in a fruitful way: not only he has won the interest of several young computer scientists in computational and theoretical linguistics, thus helping to establish this field as one of the curriculum specialities at this Faculty, but also offered a „shelter“ and research environment to those whose political background was not „reliable“ enough to apply for admission at an ideologically oriented Faculty of Philosophy but whose skills enabled them to be admitted to a less „watched“ Faculty of Mathematics and Physics. It is symptomatic for the atmosphere of that time and for Sgall's sharp eyes and good intuitions that most of these former students belong now to promising researchers and university teachers at both of the Faculties.

The other fundamental issue Sgall has been recently concentrating on is the relation of the core of language and its periphery (1, 2). These notions are also rooted in the Prague School tradition, but Sgall puts them into a broader and more complex perspective. He claims that since language is more stable in its core, regularities in language should be searched for first in this core; only then it is possible to penetrate into the subtleties and irregularities of the periphery. The relatively simple pattern of the core of language (in Sgall's view, not far from the transparent pattern of propositional calculus) makes it possible for children to learn the regularities of their mother tongue. The freedom of language offers space for the flexibility of the periphery.

Petr Sgall gives an impression of a most serious, matter-of-fact and sober person. To document that he understands good and intelligent humour and that he is creative also in this respect, we include in the present volume his „Mourphology“ paper (10) as a kind of delicatessen.

Parts B and C focus on two fundamental pillars of Sgall's linguistic theory: underlying dependency syntax (Part B) and information structure (topic-focus articulation) as a basic aspect of the sentence (Part C).

Section B (**Syntax**) contains papers extending and examining the main issues of the Functional Generative Description (FGD), proposed by the author in the 1960s, (11), (12), (13). The papers chosen for this section present the author's argumentation for the importance of the difference between linguistic meaning and ontological content, which delimits the opposition of language as a system and the domain of cognition. P. Sgall demonstrates in (13) that this distinction, known since F. de Saussure and L. Hjelmslev (with linguistic meaning characterized as "form of content"), can be determined with the help of operational and testable criteria. On such a basis, the "deep cases" (case roles, i.e. the underlying, tectogrammatical syntactic relations) can be specified as belonging to the language patterning and differentiated from a conceptualization of the scenes more clearly than with many other approaches, including that of Ch. Fillmore. Strict synonymy is understood as a condition of tectogrammatical identity. Open questions (more or less directly connected with empirical studies of texts and corpora), remaining in the specification of the list of arguments (participants) and adjuncts, are discussed in (12), where also relations other than dependency are investigated. Sgall points out the possibility to linearize even rather complex more-dimensional graphs representing projective tectogrammatical structures (including coordination and apposition) into relatively simple strings of complex symbols with a single kind of parentheses. He claims that this type of structure comes close to elementary logic and thus documents that the core of language exhibits a pattern based on general human mental capacities, which might be useful in analyzing the acquisition of the mother tongue by children. The author's subtle sense for the development of linguistic research is reflected by his participation in conceiving and constructing the Prague Dependency Treebank, a syntactically annotated part

of the Czech National Corpus. P. Sgall describes the main issues of the procedure of the syntactic annotation based on FGD in (11). Examples of tectogrammatical tree structures are given here and an outlook for the future extension of the automatic part of the procedure is discussed.

One of the most innovative contributions of Petr Sgall to theoretical and formal linguistics is his claim that the **topic-focus articulation** (TFA, Part C, see also (4)) of the sentence is semantically relevant and constitutes the basic sentence structure essential for the semantic interpretation of the sentence. As discussed now in Hajičová and Sgall (in prep.) more explicitly than before, this dichotomy is considered to be more fundamental than the subject–predicate structure of traditional grammar and of the “mainstream” theories (be it analyzed in terms of constituents or of dependency syntax). Sgall refers back to Aristotelian original understanding of ‘subject’ as ‘given by the circumstances’ (τὸ ὑποκείμενον – translated in Gemoll’s 1908 dictionary as *die gegebenen Verhältnisse* ‘the given circumstances’) and ‘predicate’ (τὸ κατηγορούμενον – *das Ausgesagte* ‘the enounced’) as what is ‘predicated’ about the ‘subject’, emphasizing the aboutness relation. It is in this sense that the content of an utterance (i.e. of a sentence occurrence) can be properly seen in the interactive perspective, as an operation on the hearer’s memory state. It should be noticed that the first paper by Sgall on TFA and its inclusion into a generative description of language was published as early as in 1967 (17). The surface word order is conceived of in relation to TFA; the differences between the surface and underlying order of items of the sentence can be accounted for by a relatively small number of ‘movement’ rules. The study of issues related to the information structure of the sentence is paid a serious attention in the Prague School history, introduced there by the studies of Vilém Mathesius in the first half of last century and continued by Jan Firbas, whose approach is critically examined from the FGD viewpoint in (14). A study of these issues was given a more intensive attention by a wider linguistic community only later in the last two decades of 20th century and it is thanks to Sgall that the position of the Czech studies on the international scene has been duly specified (15) and, even more importantly, that the attention has been focussed on the basic semantic relevance of these issues (14).

Part D (**From sentence to discourse in semantics**) gives a perspective on Sgall's views on the delimitation of the language system (linguistic competence) against the domain of cognition and the process of communication. He analyzes issues going beyond the limits of the sentence – both in the ‘dimensional’ sense (extending the scope of attention to discourse) and in the sense of crossing the boundaries of the literal meaning towards the issues of reference, cognitive content and truth conditions. Well aware of the distinction between linguistic meaning and (extra-linguistic) content claimed by Praguian scholars following de Saussure, Sgall (19) analyses the notion of ‘meaning’ as present in linguistic and logical discussions and suggests to distinguish between several explicata of the concept: (a) meaning as linguistic patterning (literal meaning), (b) meaning (or sense) as literal meaning enriched by reference, which can be understood as a layer of interface between linguistic structure and the semantic(-pragmatic) interpretation of natural language, (c) meaning in the sense of structured meaning, i.e. with specifications more subtle than propositions (Lewis-type meaning), (d) meaning as intension, (e) meaning as extension, and (f) meaning as content, taking into account the context-dependence of the content of the utterance. In this paper, as well as in all other papers on the issues of meaning, especially when discussing the distinction between ambiguity and vagueness, a crucial emphasis is laid on the necessity to establish and apply operational criteria for making the relevant distinctions. Sgall's own proposal of a starting point for a description of the semantic system of a language is presented in (20) as a nine-tuple, taking into account the outer shape of the sentence described, the representation(s) of the meaning(s) of the sentence, the entities that can be referred to, the set of items activated (salient) at the given point of time of the discourse, the possible sense(s) of the utterance token with the given meaning, the class of possible worlds, the set of truth values, and Carnapian proposition (i.e. a partial function from Sense(Meaning(Sentence)) into the class of functions from the possible worlds into the truth values). The author tests the potential of the proposed framework on several examples, each illustrating some particular point present in the discussions of natural language semantics such as the relevance of topic-focus articulation (see (4) and Part C of the volume) for semantic interpretation, the importance of the different

kinds of contexts (attitudinal, quotational) for the operational criteria for synonymy, and the cases of presupposition failure and contradictions. Discourse patterning in its dynamic perspective based on the notion of the hierarchy of activation is discussed in detail in (18) and partly also already in (20).

The papers included in part E (**Typology of languages**) are closely connected with the author's linguistic beginnings. As a pupil of V. Skalička, the founder of the Prague School typology, Sgall develops the ideas of his teacher and supervisor in (22) and (23) (see also (1)), pointing out that each of the types of languages can be understood as based on one fundamental property, which concerns the way of expression of grammatical values: by free or affixed morphemes, by a word-final alternation (a single ending), or by word order. In (24), which is a part of Sgall's habilitation about the infinitives in the Ṛgveda, the nominal and verbal characteristics of infinitive in agglutinative and inflectional languages are analyzed. While in languages of the former type the role of the "second verb" in a sentence is fulfilled first of all by verbal nouns, the latter type prefers an infinitive with a single ending (without preposition), and the analytical counterpart is a subordinate clause. In (23) the author discusses various meanings in which the terms "type" and "typology" are used in contemporary linguistics, distinguishing between polysemy of a term and different views of a single object of analysis. A type differs from a class in that it is based on a cluster of properties, on their "extreme combination". Working with one fundamental property for each type and with the probabilistic implication makes it superfluous to enumerate sets of properties defining the individual types. Agglutinative and inflectional languages are compared as for their "naturalness" (*Natürlichkeit*) in (21). Although inflection, based on a single ending with many irregularities, seems less natural than agglutination from the morphemic point of view, inflection conveys a more appropriate basis for natural syntax (with cases rendering mainly arguments or theta roles, the high degree of "free" word order expressing the topic-focus articulation, and analytical prepositions occurring in the forms of adverbials). Sgall, as always, is aware that some questions examined here are far from a finite solution (e.g. the boundaries between lexical units and syntagms or between word derivation and morphemics are still open for further discussion).

The papers included in Part F (**Speaking and writing**) reflect Sgall's permanent interest in sociolinguistic issues. The situation of Czech in everyday speech is characterized by the author as code switching rather than diglossia known e.g. from the Arabic world. Following the classical functional viewpoint of the Prague Linguistic Circle, Sgall suggests that linguists should describe the actual usage of Czech (especially of its morphemics, considered to be the main source of the differences between the varieties of Czech) in different layers of communication, rather than impose prescriptions. The position of Common Czech among the varieties differs nowadays from that of the so-called interdialects. Speakers of Czech are encouraged by the author to reduce the means with a bookish flavour in their communication, because their occurrence in other than bookish contexts is one of the reasons why the Standard norm and everyday spoken Czech are quite distant. The nature of the orthographical systems using graphemes is studied in (26), where the author provides a definition of such notions as alphabet, orthography and spelling, based first of all on the relation between phonemes and graphemes. Questions about appropriateness of orthographical systems are formulated on the basis of this explicit description. Sociolinguistic issues connected with an orthographical reform are touched upon by the author as well.

It is not only the broad scope of interests and deep insights that characterize Petr Sgall as an outstanding scientific personality. His deep knowledge and clear view of linguistic (and, in a broader sense, cultural) resources and background ranging from the historical beginnings up to the present-day modern trends is in a unique balance with the originality of his own proposals and solutions. He has never fallen into the trap of black-and-white descriptions of language phenomena: he has always been aware of the restrictions given by the complexity of the described object, i.e. language, and has found a reasonable way out by distinguishing between the notions of the centre (core) of the system and those of the system's periphery. Sgall's deep insights and capability to distinguish these two aspects is documented by his contributions throughout the present volume.

Editorial Note

Due to the richness and broad scope of Petr Sgall's publication activity, it was not at all easy to select only few of his papers for this volume. In our choice, we have followed a couple of guiding principles: we have included only papers where Sgall was the only author and exclude collective papers even if often he was their main author; for a more general acquaintance with the Functional Generative Descripition, of which Sgall is the originator, and the work on Prague Dependency Treebank, in which he has played a crucial role when setting its background and general linguistic conception, we refer to Sgall et al. 1986; Hajičová, Partee and Sgall 1998; Hajič 1998; Hajič et al. 2000; 2001; Hajičová et al. 2002. As for the dates of publication, we have concentrated on recent papers published after 1990, and this is e.g. why we have included only a very small part of Sgall's important habilitation, i.e. its section about the typological nature of infinitive.

The papers in individual parts are ordered from more recent to earlier ones. Since several important contributions of Petr Sgall were intended to introduce the Praguian views and results to different linguistic circles (to those of Systemic Grammar, to the environment of functional linguistics, to that of pragmatics, and so on), the author had to repeat the basic assumptions and standpoints of FGD. In the present collection, we preserve such repetitions only in cases when they are very short or when this is necessary to make it possible for the reader to follow the development of his ideas. This criterion is most relevant for paper (1) from 2003 and papers from the earlier dates, which are valuable from the point of view of temporal priority. We do not shorten the papers if they discuss similar issues from different angles; this is e.g. the case with the passages on presupposition, allegation, synonymy. Thus, different aspects restricting the concept of (pure, strict) synonymy are discussed in (7), (13), (19) and (20). The concept gained in this way is rather narrow so that Petr Sgall more recently proposed to use the term quasi-synonymy in those cases in which two sentences or constructions, etc. differ in their truth conditions only with a specific lexical cast (as is the case e.g. in the presence of attitudinal adverbials such as *willingly* with passivization). Also the short mention of the classification of tectogrammatical units in (4) is necessary for the un-

derstanding of that paper, but more details on this classification can be found in (11). In a similar vein, the possibility of a linearization of tectogrammatical representations is mentioned in (4) and in other places, but a more detailed account is given in (11) and (18). If possible, we avoid the reduplication by shortening one of the papers; we mark such a deletion by ((...)), mostly with a reference to that of the included papers in which a more complete formulation of the given point can be found; as the case may be, the surrounding contexts are slightly adjusted. In some of such cases, also examples and footnotes have been left out, without changing the numbering of those which have remained in the text.

Apart from the small changes mentioned above, the texts are left as they were in their original form, only misprints and similar tiny omissions are being corrected.

The present collection could not have been compiled without the extremely valuable cooperation with the author himself. The Editors are also most grateful to Anna Kotěšovcová for her devoted and time-consuming technical work connected with the preparation of the electronic version of the papers, which in case of earlier contributions involved laborious and exacting scanning. The proof reading has been done very carefully by Zdeněk Kirschner.

Eva Hajičová and Jarmila Panevová
Prague, March 2005

References

This list of references contains only papers and books referred to in the Introduction. Petr Sgall's bibliography before 1986 was compiled as a gift from his colleagues at the occasion of his 60th birthday and was made available as an internal report of the Faculty of Mathematics and Physics, Charles University; the bibliographical data from later periods were published at the occasions of his birthdays in the Prague Bulletin of Mathematical Linguistics (PBML) 55, 1991, 95–98; PBML 65–66, 1996, 113–122 (bibliography 1986–1996, with a short introduction “Petr Sgall Septuagenerian”) and PBML 75, 2001, 87–91 (bibliography

1996–2000). A complete bibliography of Petr Sgall is attached at the end of this volume.

- Gemoll W. 1908. *Griechisch-deutsches Schulwörterbuch*. Vienna: Tempsky – Leipzig: Freytag.
- Hajič J. 1998. Building a syntactically annotated corpus: The Prague Dependency Treebank. In: E. Hajičová (ed.), *Issues of Valency and Meaning, Studies in Honour of Jarmila Panevová*, Prague: Karolinum, Charles University Publishing House, 106–132.
- Hajič J., Vidová-Hladká B., Böhmová A. and E. Hajičová. 2000. A Three-Level Annotation Scenario. In: A. Abeillé (ed.), *Treebanks: Building and Using Parsed Corpora*, Amsterdam: Kluwer, 103–127.
- Hajič J., Hajičová E., Pajas P., Panevová J., Sgall P. and B. Vidová-Hladká. 2001. *The Prague Dependency Treebank*. CDROM LDC2001T10. Linguistic Data Consortium, Philadelphia. University of Pennsylvania.
- Hajičová E., Panevová J. and P. Sgall. 2002. K nové úrovni bohemistické práce: Využití anotovaného korpusu [Towards a new level of work in the study of Czech: Working with an annotated corpus] I. *Slovo a slovesnost* 63: 161–177, II. *Slovo a slovesnost* 63: 241–262.
- Hajičová E., Partee B. and P. Sgall. 1998. *Topic-focus Articulation, Tripartite Structures, and Semantic Content*, Amsterdam: Kluwer
- Hajičová E. and P. Sgall (in prep.). The fundamental significance of Information Structure. To appear in C. Caffi, H. Haberland et al. (eds.): *Future Prospects of Pragmatics*.
- Mey J., ed. 1986. *Language and Discourse: Test and Protest. A Festschrift for Petr Sgall*. John Benjamins Publ. Company, Amsterdam-Philadelphia, 611 pp.
- Sgall P. 1958. *Vývoj flexe v indoevropských jazycích, zejména v češtině a v angličtině* [The development of inflection in Indo-European languages]. Prague: Rozpravy ČSAV.
- Sgall P. 1958. *Die Infinitive im Rgveda*. AUC-Philologica 2–3: 137–268.
- Sgall P. 1964. Zur Frage der Ebenen im Sprachsystem. In: *Travaux linguistiques de Prague* 1: 95–106
- Sgall P. 1967. *Generativní popis jazyka a česká deklinace* [Generative Description of Language and Czech Declension], Prague: Academia.
- Sgall P. 1978. Three American volumes connected with Czech linguistics. *Prague Bulletin of Mathematical Linguistics* 30: 61–68.
- Sgall P., Hajičová E. and J. Panevová. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht: Reidel, Prague: Academia.

A.

GENERAL AND THEORETICAL ISSUES

1.

TYPES OF LANGUAGES AND THE SIMPLE PATTERN OF THE CORE OF LANGUAGE

1. Introduction

I would like to present and substantiate several points which may be of crucial importance for the development of theoretical linguistics, although they have been largely neglected in the “mainstream” trends. After several decades of investigations and discussions first in the context of V. Skalička’s typology and then in that of the Functional Generative Description, I am convinced that they are significant:

1. The *physical* character of natural languages, which use phonetic means, brings about *strict limitations* on the relationships between meaning and expression. The limitations are decisive for the ways how grammatical values are conveyed. Basically, the grammatical values can only be expressed by (a) morphemes (b) alternations, and (c) the order of lexical items in a sentence. These properties constitute the fundamental background of the types of languages. A characterization of the types is discussed in Section 3.1 below, after a brief examination of the nature of general concepts used in connection with language typology in Section 2.

2. One of the main results of the history of typological thought is the transition from overestimation of Indo-European (old or modern) languages to the recognition that typological change does not constitute a line of “progress” and that language types are *not* directly connected to the semantic *richness* of languages (see Section 3.2 below).

3. The interactive nature of language is reflected in the sentence structure by the opposition of topic (T) and focus (F), i.e. by the relation of *aboutness*, of a “psychological” predicate F and its argument T as the

background of the meaning (i.e. of the underlying structure) of the sentence, with T being the linguistic counterpart of “given” and F that of “new” information. In the general case, a sentence cannot be interpreted on the basis of a predicate structure of a shape similar to $R(a_1, a_2, \dots, a_n)$, with the arguments a_i in the positions of subject, objects, etc., but only as $F(T)$, with its negative counterpart corresponding to $\neg F(T)$, i.e. adding an operator (negative, positive), and working with typed lambda calculus. The T-F articulation, the analysis of which is outside the scope of this paper, has been characterized as one of the basic aspects of sentence syntax in Sgall et al. (1986), Hajičová et al. (1998).

4. Distinguishing between (underlying) *sentence structure* and the *morphemic* means of its expression in the sense of paragraph 1, it is possible to handle the core of the language system as patterned in a relatively simple way (using a dependency based grammar and an underlying order of T preceding F), see Section 4. In this way it may be seen that the pattern of the core of language comes close to what usually is supposed to belong to the general human *mental abilities*; the large and complex periphery of language can be described as consisting of contextually restricted specific deviations from the core (see Section 5).

2. General concepts in typology

2.1. Terms and notions

It is crucial for our discussion to distinguish between (a) different meanings of an ambiguous term, and (b) different approaches to a single object of study. While in case (a) it is relevant to ask about the terminological appropriateness of the different uses of the term, in case (b) the question is which of the approaches is more adequate to the given object. Thus, one should distinguish:

(i) whether the term ‘typology’ is used in the sense of a theory of *language types*, or without being connected with a notion of type; in the present contribution, language types are involved, so that attention is not devoted to partial typologies, such as those oriented towards the semantics

of quantification (as with B. H. Partee), incorporation (E. Bach), or other more or less specific layers (as e.g. W. Croft's or T. Shopen's writings on syntax), or towards areal typology,¹

(ii) whether different uses of the term 'language typology' concern a single object or not, i.e. whether the investigations aim at an analysis of languages covering their basic properties; this question certainly is not simple, since the use of the term *language type* varies from one author to another in many aspects, some of which are discussed below.

In any case, these distinctions (cf. Sgall 1971) are relevant for a delimitation of different aims, each of which certainly is interesting and important in its own right.

2.2. The concept of type of language

A methodological basis for typology may be found in Hempel and Oppenheim (1936), who use *comparative* concepts and characterize objects as more or less close to a given point (or extreme).² The notion of the type of language was developed in a similar orientation in the classical *Prague Linguistic Circle*, in which Skalička (1935; 1979) characterized the type of language as a collection (cluster) of properties intrinsically connected by *probability implications* of the form: "if a language has the property A, then it probably also has the property B", i.e. $P(A,B)$. The fundamental nature of these probability implications was specified already by G. von der Gabelentz (1894, 5f), according to whom "...die Erscheinung A trifft mit so und so großer Wahrscheinlichkeit mit B, C, D usw. zusammen, selten mit E, nie mit F." Skalička assumes that the probability implication P is a symmetric relation: if $P(A,B)$ holds for a pair A,B, then $P(B,A)$ also holds.

From the understanding of the notion of type as based on probability it follows that the types are ideal extremes not fully attainable by existing languages; the languages come closer or less close to one (or more) of the existing types, with properties of different types cooccurring in the structure of every existing language. Typology, based on clusters of properties, thus differs from a classification, since the latter can be based on a single

property and distinguishes between classes on a yes/no basis, i.e. strict *classification* is based on a partition of a set. Let us just recall that language typology and the research in language universals – known esp. from H. Seiler – study similar classes of phenomena, although they analyze problems of different kind.

3. Types of languages

3.1. Language types as based on the means of expression of grammatical values

If the concept of type of language is understood along the lines discussed in Section 2 above, then it appears as appropriate instead of Skalička's "mutual" favorability of the typologically relevant properties to work with their favorability as an asymmetric binary relation, i.e., from $P(A,B)$ neither $P(B,A)$, nor $\text{non-}P(B,A)$ follows. With this approach, the Praguian typology has been brought to a stage at which it is not necessary to work with lists of typologically significant language properties, as Skalička did. It is now possible to identify a single *property* as favorable to all the other features characteristic of a certain type (see Sgall 1995 and the writings quoted there).

Such a fundamental property may be seen in the way of expression of grammatical values, understood in a broader sense, as including the formation of lexical units, i.e. derivation of words, their composition, borrowing from other languages, or creation of lexical units composed from more than one word. While lexical values are conveyed by strings of phonemes (lexical morphs, roots) in all languages, grammatical values have different means of expression, the repertoire of which is limited by the conditions given by the phonic nature of language.

It is then possible to work with three instances of the fundamental property as giving rise to five types, thus arriving at the types which have been specified in the 1930s by V. Skalička on the basis of the tradition of typological research (having culminated in the works of F. N. Finck and of E. Sapir) and of his extensive knowledge of many languages from all parts of the world:³

(a) The grammatical values are also expressed by morphemes, which themselves are rendered by *morphs* (strings of phonemes), and these can have one of two different forms:

(a1) they resemble the strings conveying lexical meanings – both kinds of morphs often are monosyllabic, and their word order positions vary at least within certain grammatically fixed boundaries; this is the *analytic* type, in which grammatical values are conveyed by function words; this basic property is favorable to (i) an abundance of dependent (embedded) verb clauses with conjunctions, (ii) the absence both of case endings and of agreement, which is favorable to grammaticalized word order (i.e. to the main syntactic functions expressed by positions of the words, e.g. with the order SVO required by grammar); the word order cannot then be used on a large scale to distinguish topic and focus; this is favorable to the presence of articles (since definiteness has much in common with givenness and indefiniteness with introducing new discourse referents); the absence of affixal derivation is favorable to lexical conversion (e.g. the type *stone wall* in English) and to numerous loan words; many words are unmotivated, non-derived, and may then well be monosyllabic, which is favorable to a large number of vowels, being useful to distinguish the short words; English, French, but also e.g. Hawaiian are Skalička's examples of languages coming close to the extreme of this type, although each of them is far from reaching such an extreme;⁴

(a2) or the grammatical and derivational morphs differ from lexical morphs in being attached to them as affixes each of which expresses a value of a single category; this is the type of *agglutination*, with a high number of morphemic cases (although the subject is expressed just by the bare lexical morph) and an unclear boundary between them and adverbs derived from nouns (i.e. between grammatical morphemics and lexical derivation), with many deverbal nominals (nouns of action, of actor, of artefact, adjectival participles, etc.), long word forms, and (since in such word forms the differences between all the phonemes are not necessary to distinguish the forms) with phonemic reductions such as vowel harmony; languages close to this type are e.g. Turkish, Hungarian, Finnish, Georgian, Basque, Eskimo, Armenian.

(b) The grammatical values are expressed by modifications of lexical morphs, i.e. by *alternations*, which occur either

(b1) at the end of the lexical morphs, as is the case with the *inflectional* (fusional) type: a single ending is present with every word form, with no clear boundary line between the ending (i.e. alternation) and the stem, so that we face an alternation of the end of the lexical morph, rather than an affix (e.g. Lat. *anima, animae, animis*); the single ending expresses a set of cumulated functions (case, number and gender, or person, number, tense, mood and diathesis, etc.); the case of the subject (nominative) has its own endings; the endings exhibit a high degree of synonymy and of ambiguity; agreement abounds (a verb agrees with its subject, an adjective with its head noun); a high degree of “free” word order is present, which (together with the sentence prosody) conveys the topic-focus articulation; a dependent (embedded) verb often has the form of an infinitive without a preposition; inflectional conversion is frequent (i.e. derivation has the form of subsuming the derived word into another class than the derivational basis (Lat. *anima – animus*); inflection prevails in Latin and other Old Indo-European languages, in Russian, Czech, etc.;

(b2) or in the inner part of words – *introflexion*: phonemes (especially vowels) occurring within individual lexical morphs serve to express word formation and morphemics; this basic property is favorable to further features similar to those of inflection; in Semitic languages this property is combined with agglutination.

(c) Grammatical values are expressed just by the *order of lexical morphs – polysynthesis*: the boundary between lexical and grammatical units is unclear, there are many compound words (Vietnamese, written Chinese, Yoruba, Thai).⁵

Even from this very short survey it may be seen that the five language types do not concern only the morphemic shape of word forms. On the contrary, the strength of the different types in individual languages is relevant for many layers of the language systems and for many features of communicative activities. Also *syntactic* properties come into play, such as the form of a dependent verb or of subject and object, the presence of agreement, or the presence of a copula vs. that of an agglutinative affix at the predicate nominal (as in Turk languages, see Giger and Vykypěl 2001). In phonemics, e.g. the number of vowels (and the presence or absence of their harmony) is involved; cf. also Plank (1998), who, quot-

ing G. Fenk-Oczlon and A. Fenk, writes that agglutination is favorable to simple syllable structure.

Thus, Comrie's (1981: 78) requirement according to which language types should cover most different layers of language appears to be met to a relatively high degree by this approach to typology. In any case, a fully holistic typology seems to be excluded; it is an empirical fact that the clusters of typologically relevant properties never have been found to be strong enough to cover the whole of a real language, rather than just as theoretical, idealized constructs (cf. Note 4, as for the disadvantages which would concern the "complete" types).

The advantages of implication laws stated by Holenstein (1985) are preserved in the probabilistic approach. This concerns above all the possibility to identify a *hierarchy* among the properties concerned, which is lost if only a prototype and its periphery are distinguished (and also if one works with symmetrical 'favorability', which made Skalička to formulate lists of properties of individual types, rather than to adduce their fundamental properties). The basis of this hierarchy is anchored in the opposition of lexicon and grammar, i.e. in a fundamental opposition present in every natural language (although the boundary line between these two domains is not clearcut, especially with respect to the opposition of cases and denominal adverbs, to that of verbs and deverbal adjectives, or to certain pronouns as opposed to personal endings).

The fact that the implication laws underlying typology are based on probability makes it necessary to work with a *quantitative* evaluation. It is important to base this evaluation on values that are of fundamental importance for the language types, rather than on those that are easily accessible for counting. Thus, it is not sufficient to concentrate on the length of sentences or of word forms. Instead of this, e.g. a procedure allowing for the identification of the degrees of inflection may take into account phenomena decisive for this type, such as the cumulation of functions in the ending, or the synonymy and the ambiguity of endings. Using such a procedure (see Sgall 1983b), a characterization of Czech declension resulted in the following values: 58 morphemes, 116 endings, 26 values of morphological categories, 34 morphs; three relevant indices can then be assigned the following values:

index of cumulation of functions – $58 : 26 = 2.23$,
index of synonymy of endings – $116 : 58 = 2$,
index of ambiguity of endings – $116 : 34 = 3.41$.

This method was applied (with a certain simplification) for a comparison of Slavic languages by I. I. Revzin and his colleagues, see Volockaja et al. (1963). Also further studies devoted to a typological comparison of languages on the basis of Skalička's view, which check and enrich this approach, confirm that every language contains properties of different types, which is a consequence of the probabilistic character of Praguian typology, and one of its basic ingredients.⁶

If the 'natural morphology' approach of W. Mayerthaler and others (now see Dressler 2003) is taken into account, it is possible to see at least a certain degree of naturalness (especially of iconicity) in some of such combinations of different types (cf. esp. Popela 1991, 1999, Nau 2001, Giger and Vykypl 2001, Sgall 1988b). Thus, agglutinative features in word derivation often occur in languages of other types thanks to the prototypical situation in which a semantically specific derivation base is combined with affixes having more general meanings (e.g. diminutives, feminine nouns derived from masculines, or J. Kuryłowicz's 'syntactic derivation' switching the parts of speech). On the other hand, inflectional morphemics (even with its irregularities) seems relatively appropriate for a short way of expression of the most frequent kinds of word forms, i.e. especially for the case forms that express subject and objects (or Actor, Objective, Addressee, etc.). The other (adverbial) cases often are connected with prepositions, i.e. analytic function words, in inflectional languages as well, and such means seem appropriate to express adverbial and attributive relations between autosemantic lexical words; they connect two such words, whereas a derivational affix just expresses a semantic adaptation of a single word.⁷

Many other phenomena have features of two or more different types, such as the ergative sentence structure, or suppletive forms. Often even individual word forms exhibit properties of more than one type, e.g. those of analysis and of inflection in combinations of prepositions and articles such as French *du*, *des*, *au*, etc., German *am*, *vom*, *zum*, not to speak of inflected analytic function words such as articles distinguish-

ing gender and number in the just cited languages (Schwegler 1990, 148) or of the affinity between function words and affixes (Plungian 2001), both of which represent grammatical means having the form of morphs, i.e. belong to the class of types specified as (a) in Section 3.1 above.

The importance of typological studies for understanding issues of diachrony has been known for more than a hundred years. Gabelentz (1901: 255–258) characterized the typological development of languages as a spiral going from “isolation” (or, in Skalička’s terminology, polysynthesis) to agglutination, from there to inflection, then to new “isolation” (or analysis). This movement, which differs from a trend of enriching the lexicon and the grammar, can be understood as a more or less regular rotation of types, rather than as a route for the “progress” of language. The existence of exceptions to Gabelentz’s spiral was pointed out in Skalička’s (1941) observations on the declension in the eastern branches of Indo-European languages, which develop from inflection “back” to agglutination. Perhaps this is due to the rapid development of civilization leading to higher regularity, and to different internal conditions for such external factors: if in the epoch of rapid external changes inflectional endings were weakened more or less in a language, this might have been decisive for the growth of regularity to reduce either the endings themselves, substituting them by prepositions, or the number of different paradigms and other inflectional intricacies.⁸ The typological development of the Indo-European languages was properly enriched by Vennemann’s (1974) analysis of the topic-focus articulation.⁹

The Prague image of the types of languages may be understood on the one hand as a result having its roots in the history of linguistics, to which we turn now, and, on the other hand, as offering certain highly important insights into the nature of language (see Section 4 below).

3.2. Three main lessons from the history of typology

Only an extremely brief comment on certain points of the development of linguistic typology can be presented here, based on more detailed studies (summarized in Sgall 1995) and intended to help de-

limit the position of Prague typology among the trends studying types of languages.

As Ramat (1995) recalls, in the 18th century G. Girard distinguished between “analogous” and “transpositive” languages; this may be compared to the difference between languages with a fixed SVO word order and those with a higher degree of “free” word order (and with rich sets of inflectional endings). This forerunner of typology, and also his successor A. Smith, understood the language types as given once and for all, recognizing a possible change of type only in cases of “mixing” of types. Similarly, in A. W. Schlegel’s approach one of the basic ideas was the absence of a change of type. The view of language types as rigid categories connected to some psychological values of ethnic groups was more or less clearly discarded by Humboldt (1836),¹⁰ but emerged again, especially with H. Steinthal’s ethnopsychological view of Indo-European languages as superior. Another turn towards a sober analysis of linguistic structures themselves, including the interplay of typologically different properties within a single language, can be found in the adaptation of H. Steinthal’s work by Misteli (1893). Also the works of Max Müller, A. F. Pott, A. Hovelacque, W. D. Whitney and other linguists of those epochs are highly illustrative from this point of view. Later, F. N. Finck changed his attitude along these lines between his two books (1899, 1910).

A similar change can be seen in the reasoning on diachronic typology – from A. Schleicher’s “Sprachaufbau-” and “Sprachzerstörungsperiode” with the Proto-Indo-European in between as an ideal stage, exhibiting all the richness of morphemics, and Jespersen’s (1894) “progress in language” (with Modern English as superior to other languages) to G. von der Gabelentz’s “Spirallauf”, mentioned in Section 3.1 above.¹¹ Gabelentz’s expectations concerning what now could be called a holistic typology can be characterized by his (1901, 481) requirement that our knowledge of any relevant property of a language should immediately lead to a specification of its other properties, similarly as in biology knowing a leaf we know the properties of a tree (see Plank 1991).¹²

E. Sapir’s approach was fully freed from the old prejudices, but lacked a systematic attention to the interconnections between individ-

ual linguistic properties. Such interconnections were studied in Prague especially in Mathesius' (1928) 'linguistic characterology' and in connection with Jakobson's (1929; 1958) revealing idea of implication laws. Jakobson pointed out the goal oriented, teleological nature of language (now see Leška 1986; 1987, Toman 1995, p. 141, and Sériot 1999a).¹³ Within this linguistic context Skalička (who understood laws of the form of strict implications to be important for phonology, rather than for grammar in the narrow sense) specified his five types as based on probabilistic affinities and constituting a relatively highly holistic characterization of languages. The just quoted formulation of G. von der Gabelentz may find at least its partial parallels in Prague typology: e.g., if we know that a given language typically expresses a syntactic connection of two nouns by (i) a preposition, (ii) an ending, (iii) an affix, or (iv) a compound, then we can predict this language to exhibit other properties of the relevant type, too, i.e. (i) analysis (having embedded clauses introduced by conjunctions, an infinitive with a preposition, articles, many auxiliaries and monosyllabic words, etc.), (ii) inflection (with a complex set of personal and case endings, a prepositionless infinitive, agreement, a high degree of "free" word order, and so on), (iii) agglutination, or (iv) polysynthesis.

The predictive power of Praguian typology is certainly lower than ideal, due to the probabilistic nature of this approach. In another respect, the predictive power depends on the scope of the clusters of properties which constitute the types. The clusters are larger with the Praguian approach than with the word-order based typology of Greenberg and others (see esp. Mallinson and Blake 1981), which has brought many highly valuable insights not only into the word order, but also into the general conditions of the order of morphemes within words in most different languages. Let us just remark that the differences of the degrees of "free" word order might be assigned much more importance for the characteristics of language structure than is done with the just quoted approach. The most frequent word order in Latin (Czech, Russian, etc.) is SVO, as is the case in English, French, and also in Chinese; however, it does not seem to be optimal to regard all these languages as belonging to the same type. Greenberg (1995) opens a way to a much more general view of typology, especially to that of the change of type.

In another sense, search for a holistic typology may be seen in Coseriu's (1980; 1983) approach, which stresses the integrational character of typology; however, diachronic transitions between types are not abrupt and thus phenomena of different types are present in a single language, as Coseriu (1980, 169) admits, being aware of the difference between type and class (p. 167). The large set of writings of Coseriu and of his followers certainly is the richest source of European typological thinking in general and especially of Romance linguistics. However, it still belongs to questions open for further discussion what is the degree of the explanatory and predictive power of his integrational typology, which understands a type primarily as being characteristic of a single language as the basic patterning of its functional layer. A comparison with Skalička's approach has been presented by Geckeler (1988), see also Dezső (2000) and Kretz (in prep.).

From regarding the types of languages as rigid categories connected to assumed psychological values of ethnic groups, the development of research has led to a sober analysis of linguistic structures themselves, bringing to the foreground especially the following points:

(a) Large-scale differences between the structures of languages are determined by relationships displaying certain degrees of probability, which constitute clusters of properties, language *types*.

(b) The image of language types should not be blurred by their assumed connections with some intrinsic psychological values of languages or with their semantic richness. As we know from Skalička and his predecessors, the types are basically semantically *equivalent*. The experience corroborates the view that although translating between English and Czech (or Japanese, Chinese, Arabic, and so on) certainly is a more complex task than translating between English and French, still the typological differences only make the translation process less easy rather than impossible.

(c) The typological differences concern relationships between (*underlying*) syntax and (*morphemic*) surface, as was stated by Ramat (1985, 20); more specifically, they are based on the way of expression of grammatical values (and of word formation), see Section 3.1 above.

4. Typology and the nature of language

4.1. Typology as a challenge for theoretical linguistics

A theoretical descriptive framework should allow us to describe a language as preferring a certain type. Close connections between parameters, especially those concerning the relationships between (underlying) sentence structure and morphemics should be reflected. Stratificational models distinguish the defining functions of transducing automata (suitable e.g. for movement rules) from matrices of locally conditioned modifications (such as the choice of morphs); the latter are connected with a 'lower cost'. Perhaps different descriptive frameworks can correspond to different language types, with properties of other types being marked, 'more expensive'. The teleonomic explanations of the favorability among properties should be analyzed from the viewpoints appropriate for goal-directed systems.

4.2. Fundamental oppositions within the language system

If (as just mentioned), the differences between the types of languages are determined by relationships between sentence structure itself and its means of expression, i.e. between the levels of Curry's (1962) tectogrammatics and phenogrammatics, then it is important for understanding the nature of language to be aware that these relationships belong to the essential properties of its structure. An analysis of the opposition of these two levels has been carried out in the descriptive framework of Functional Generative Description, FGD, see Sgall et al. (1986), Panevová (1994), Hajičová et al. (1998). FGD works (along with phonemics and phonetics) with two sets of sentence representations, viz. with the tectogrammatical (underlying syntactic) level and with a level of morphemics. In the tectogrammatical representations (TRs), the lexical occurrences proper (autosemantic) are represented by nodes in dependency trees (or, with the inclusion of the relations of coordination and apposition, in more-dimensional networks with a basically simple patterning).¹⁴ The tectogrammatical correlates of function words have the form of indi-

ces of the lexical node labels, i.e. of syntactic and morphological symbols representing values of functors (Actor, Objective, Addressee, Manner, Locative, Cause, etc.) and of grammemes (Plural, Preterite, Imperative, Comparative, etc.), respectively. While one of the dimensions of the tree (schematically indicated as “top-down”) corresponds to syntactic dependency, the other dimension (left-to-right) serves to represent the underlying word order (the scale of communicative dynamism in the terminology of the topic-focus-articulation theory). On the other hand, a morphemic sentence representation has a single dimension (left-to-right, surface word order), it is a string of more or less closely connected symbols, i.e. morphemes.¹⁵

What we face here is the opposition between (underlying) syntax and (morphemic) surface. It appears to be appropriate not to continue working with an intervening level of “surface syntax” in the theoretical description of language (see Sgall 1992). Thus, essentially, we are coming back to the classical opposition between *syntax* and *morphemics*. Similarly to that between *grammar* and *lexicon*, this is a pair of old concepts, richly discussed and shown as a cornerstone of plausible hypotheses. Both these oppositions have served for many centuries of linguistic research, giving ground for a modular understanding of language, although their boundary lines are blurred by “grey zones” of transitions, of intermediate phenomena (between morphemics and word formation, syntax and analytic morphemics, in idiomatics, and so on).

4.3. The simple pattern of the core of language and its vast periphery

In the Prague School, the dichotomy of the center and the periphery of the language system always has been treated as one of the most important oppositions. Its analysis has been founded on R. Jakobson’s concept of markedness, the hermeneutic and theoretical roles of which are stressed esp. by Battistella (1995).¹⁶ It is impossible to overlook the numerous aspects this concept in its different forms and sectors shares with prototype theory. What is especially relevant for us is the relationship between the relatively simple pattern of the unmarked phenomena, determining the core of language, which can be captured as based on a *simple pattern*, coming close to systems that

may be understood as innate on independent reasons (propositional calculus). This view is made possible if the unmarked layer of sentence structure is accounted for by dependency trees with complex node labels, and unmarked, regular relations between syntax and morphemics are seen as prototypical.

Even if, along with different relations of syntactic dependency, also coordination and the topic-focus articulation are taken into account as aspects of sentence structure (of the TRs), the patterning of sentences can be described as (more-dimensional) tree-like objects that may be univocally represented just by *bracketted strings* of symbols, each of which comprises a lexical component and indices for the values of grammatemes and functors, cf. e.g. the (simplified) representation (1') for the sentence (1):

(1) Jim and Jane's son, who were present there, belong to the BEST specialists.

(1') ((Jim ((Jane)_{Appurt} son_{Sing.Def}))_{Conj} (Restr (who_{Plur})_{Act} (there)_{Loc} be_{Pret.Decl.} Imperf (Obj present)))_{Act} belong_{Pres.Decl.Imperf} (Dir specialist_{Plur.Def} (Restr good_{Superl})))

Note: Every dependent item or collocation is enclosed in its pair of parentheses, the indices of which, i.e. functors, denote (i) either a dependency relation with its index attached to that parenthesis that is oriented towards its head: Appurt(enance, broader than Possession), Restr(ictive Adjunct), Act(or), Obj(ective), Loc(ative), Dir(ectional), etc., (ii) or a coordination construction with its symbol attached to the right parenthesis: Conj(unction), Disj(unction), etc. The indices at the lexical item (indicated only by its orthographic form, which has to be substituted by a symbol for lexical meaning), i.e. grammatemes, correspond to the morphological values: Sing(ular), Def(inite), Pret(erite), Decl(arative), Imperf(ective), Superl(ative), etc. The items written to the left of their heads are contextually bound (in topic, in the prototypical case), those to the right of their heads are non-bound (in focus).

The transition between TRs and the surface (*morphemic*) forms of sentences can be handled by a set of rules (including movements) that does not surpass the generative power of one or two (subsequent) push-

down transducers, so that the whole description of language is not much stronger than context-free (cf. Plátek and Sgall 1978). A highly significant task then is to specify different forms of such devices appropriate for languages preferring one of the types (i.e. connected with a lower cost for the chosen type).¹⁷

Non-prototypical, marked phenomena in language are responsible for the existence of a vast and complex *periphery* of its system. Three layers of marked, secondary phenomena may be distinguished:

(a) marked members of grammatical, semantically relevant oppositions within the language core, such as the morphological values of Plural, Pret-erite, Imperative, or the contextually bound items, and so on;

(b) peripheral phenomena in the TRs, which constitute the marked layers of underlying sentence structure, e.g. coordination and apposition, or marked positions of focusing operators (see Hajičová et al. 1998), i.e. phenomena which require a more complex set of rules (or of descriptive devices) for specifying the set of TRs;

(c) contextually restricted relations between TRs and morphemic (or phonemic and phonetic) representations of sentences, which constitute a very large domain, ranging from ambiguous and synonymous items in the lexicon and in morphemics (with the sets of inflectional paradigms, their irregularities, etc.) to instances of surface word order not corresponding directly to the scale of communicative dynamism.

The core of language with its relatively simple structure is substantial for the child's acquisition of language; on the other hand, the complex, large periphery can be mastered by children step by step, with the specific, contextually restricted deviations and exceptions internalized one after the other, on the basis of analogy. Also a theoretical description capturing the core of language by relatively weak means (equivalent to a context-free grammar) perhaps should be accompanied by models of the non-prototypical subdomains and exceptional phenomena, based on lists of items relevant for the contextual restrictions of the marked points. Such a description perhaps may be based on an alternative mathematical approach working with the concept of a collection or semiset, in which the set membership is not fully delimited, see e.g. Vopěnka (1989).

Such a description should reflect the presence of two fundamental tendencies:

(a) the *consistency* of the *type* that is basic for the language described and the properties of which should be connected with low cost in the description of the relation between the underlying and the morphemic levels of this language (the low cost might also correspond to a high degree of communicative efficiency), and

(b) *naturalness* in the sense mentioned in Section 3.1 above, which might underlie a mechanism that would restrict the extreme type.¹⁸

Thus, the concept of *markedness* may play an important role also in connection with typology: within the structure of a language, the consistency of a given type can be understood as the unmarked, prototypical case, and one of the main tasks is to specify mechanisms appropriate for the description of the co-existing properties of other types, i.e. deviations, marked cases.

If the methodological requirements brought in with the Chomskyan revolution, i.e. explicitness in linguistic thought, not just in description, are to be connected with the main results of European functional and structural sources, then a simple pattern of the core of language can be gained: a set of TRs in the form of dependency trees, the possibility of a linearization of which (and of more-dimensional networks including coordination) documents its fundamental perspicuity, see (1') above. A formal description of this core then may start from a view of prototypical relationships between the TRs and the morphemic strings. Only an extremely simplified scheme may be presented here as a starting point:

tectogramatical word form: $\text{root}, (d_1, \dots, d_n), g_1, \dots, g_m, c$

d – derivational suffix (prefixes are to be handled similarly)

g – grammateme value (value of a morphological category)

c – functor value (syntactic relation)

n, m – natural numbers

Agglutination in its extreme form is based on a one-to-one relation between the TR and the means of its expression (morphemes), cf. e.g. Turkish *ev-ler-im-iz-in* 'of our houses' with the root *ev-*, d_1 *-ler-* 'Plur.', d_2 *-im-* '1st Pers. Possess.', d_3 *-iz-* 'Plur. Possess.', c *-in* 'Genitive'.

Analysis: function words correspond to the functors and grammatememes, and no suffixes are present, cf. e.g. E. *to a house of mine*.

Inflection: the functors and grammatememes are expressed by a single ending (more precisely, by an alternation of the stem at its end, or, with *introflexion*, in the middle of it), e.g. Czech *matka* ‘mother.Nom.Sing’, *matce* ‘mother.Dat.Sing’, *matek* ‘mother.Gen.Plur’.

Polysynthesis: the most frequent functors are expressed by the order of bare word roots (e.g. in a pattern such as SVO or SOV), the expression means of other functors and of the grammatememes belong to other types or are of an intermediate character (words having a rather general meaning may be ambiguous, used also in grammatical functions, e.g. an equivalent of the verb ‘give’ in the function of Dative).

Many questions of different kinds remain open, among which one of the main is how a description of the relationships between TRs and morphemics can best account for the two fundamental tendencies specified above as (a) and (b), i.e. the *consistency* of a type (connected with a low cost in the descriptive framework to be chosen) and its *limitations* (connected with naturalness and/or economy).

5. Conclusion: Typology helps understand the nature of language

We have seen that, if Skalička’s specification of the language types is adapted to the asymmetric view of favorability, then the Prague typology of languages discloses an image of the fundamentals of types as anchored in the way of expression of grammatical values. This image requires systematically to distinguish between the level of sentence structure and that of morphemics; it may be assumed that within the core of language the relations between units of these two levels are unmarked or prototypical. The pattern of the core is *relatively simple*, coming close to that of the propositional calculus and of other systems which on independent reasons may be viewed as being generally accessible to humans, i.e. determined by their innate properties. This may be useful in explaining the *easiness of acquisition* of language, analyzing the child’s language acquisition as founded on the interactivity of language in communication (cf.

Schnelle 1991), rather than on complex innate mechanism specific for the language faculty.

Acknowledgments: The work reported in this paper was carried out with the support of the project of the Czech Ministry of Education LN00A063. My personal thanks belong to F. Čermák, E. Hajičová, M. Giger and B. Kretz, who helped me with their valuable comments.

Footnotes

- ¹ Highly important studies on causation, resultative, iterative and other constructions, as well as on general issues of typology, have been presented by the group of A. Xolodovič, as summarized by Nedjalkov and Litvinov (1995). Also J. Nichols' characterization of language structures with syntactic relations marked either on the head or on the dependent word is of great interest, although often both kinds of marking co-occur, e.g., in many languages, the agreement of the verb with its subject is accompanied by the agreement of the adjective with its head noun and/or by case marking; cf. also Stolz and Urdze (2001).
- ² If Ineichen (1979, p. 2) claims that such a comparative approach is of advantage in medicine or psychology, rather than in linguistics, then it may be recalled that intermediate grey zones blurring the boundary lines between subdomains (and classes) of phenomena certainly are present in natural languages, and are crucial for their description (cf. e.g. Lehmann 1990 on preponderant classes and residues; also Sgall 2002).
- ³ Other layers of means of expression could be looked for in the domain of intonation and prosody, but these hardly can serve as something more than features accompanying the main means of expression (in the lexicon, as e.g. in Chinese, or in grammar, as e.g. in the case of interrogative intonation in several languages). Also a type (b3), with inflectional alternations at the beginning of word stems probably would lack the necessary minimum of perspicuity.
- ⁴ The extreme of the analytic type would contain no endings or affixes and no compounds, so that the lexicon of such a language would be fully scattered, without productive derivation means, cf. *sir* vs. *lady*; there would be no word class boundaries, i.e. an unlimited conversion such as that of *stone wall* vs. *wall stone*. An extremely agglutinative language would have only a single word basis, from which all lexical units would be derived by long strings of affixes. An extremely inflectional language would have a specific paradigm for every noun, adjective and verb, so that there would be no exaggeration in characterizing the language, as Skalička did jokingly for Old Greek, as a nice language the learning of which requires no tedious memorizing of vocabulary, since having learned its grammar you know its lexicon.

- ⁵ Skalička rejected the older use of the term ‘isolating’, which sometimes was connected with an evaluation of such languages as displaying only poor grammatical patterns. Therefore, he used the term ‘polysynthetic’ and spoke about the isolating type of English or French, which has no affixes or endings, preferring function words, isolated from their lexical words, and which favors unmotivated lexical units, i.e. a dispersed lexicon, cf. Note 4. In M. Giger’s introductory remarks to Skalička (2002) more data can be found concerning Skalička’s writings on the notion of incorporation, as well as on Finno-Ugric languages.
- ⁶ See esp. the analyses of Japanese (and of general linguistic issues concerning variation, communicative competence, and so on) by Neustupný (1978), of Slavic languages by Ďurovič (1973), more specifically of the West Slavic domain by Weiss (1983), Lotko (1997), and Giger (1998); Czech is compared to Russian by Popela (1988), to Baltic languages by Giger and Vykypl (2001), and the interplay of different typological properties in West-European languages is examined by Čermák (1978), Uhlř (1969;1988) and Geckeler (2001).
- ⁷ This concerns also the fact that Semitic languages are predominantly agglutinative, although they exhibit a larger amount of introflexion than that present in other languages (Rubba 2001), as well as the well known agglutinative features in Latin conjugation (see e.g. Bossong 2001, who, without offering a new solution of the issue of language type, comments on Skalička’s views in a rather superficial way, seeing indiscriminate (“pauschalisierende”) classifications in an approach that in fact works with properties of different types as combined in individual languages).
- ⁸ In his spiral, Gabelentz does not distinguish between the types he called analytic and isolating; on p. 257 he says that English seems to rush towards a pure isolating system (“...dem rein isolierenden Systeme zuzueilen scheint”). It may be assumed that while the analytic type occupies the position between inflection and agglutination in the spiral, the isolating type (called ‘polysynthetic’ by Skalička, due to its way of word formation) stays apart, perhaps being suitable as a starting point of the whole development, never to be reached again. However, Gabelentz only speaks of the spiral development itself, rather than of a fixed state of its origin, not presenting any possibility to characterize individual languages as more and less “developed” (standing less close or closer to such a state). A further question would concern the (im)possibility to investigate how many times a part of the rotating spiral has been passed by a language (cf. the “secondary” agglutinative forms that include affixes developed from pronouns in some of the Romance languages, e.g. Italian *dámelo*). Even for Indo-European languages we do not know much about their prehistoric development, although before the (partially and hypothetically) reconstructed shape of their common source a number of such cycles could have been absolved. Even less can be said about the stages that may have preceded the known systems of languages of other families, most of which nowadays are agglutinative or analytic.
- ⁹ Kurzová (1993) and Hoskovec (1999–2002) do not understand the oldest known (and reconstructed) shape of Indo-European languages as a direct witness of agglutination, and speak only of a stage of ‘derivational inflection’ as preceding that of ‘paradigmatic

inflection', the latter subtype corresponding to Skalička's concept of inflection and the former to a stage in which the word forms in the sentence were less strictly grammaticalized (with a smaller range of agreement and of determination of case by the governing lexeme, lacking a clear boundary between lexical derivation and morphemics). It remains to be discussed whether this stage can be interpreted as a transition from an even older agglutinative system to (paradigmatic) inflection proper.

- ¹⁰ The beginnings of typology are described by Dezső (1999); many new insights on different approaches are offered by Kretz (in prep.).
- ¹¹ A view similar to that underlying Gabelentz's spiral, as well as its synchronic counterpart, was characterized by Lehmann (1985b) as a scale of grammaticalization, which comes close to Prague typology in several respects. Another branch of grammaticalization can be seen in changes such as those mentioned in Note 9. Also cases of degrammaticalization have been registered, see Kim (2001).
- ¹² As for the question of terminology raised by Plank (1991), it still seems possible to believe with Ineichen (1991, p. 1) and others that Gabelentz wanted to speak of typology, although the editors used the mistaken label "hypology" in the title of his posthumous paper, as well as then (perhaps without an effective checking) in the table of contents.
- ¹³ In Romance studies, this nature of language is discussed e.g. by Schwegler (1990, 177–183), quoting A. Martinet and H. Geisler.
- ¹⁴ Dependency based syntax, known in European linguistics since the 1830s (thanks to K. F. Becker) and systematically elaborated by L. Tesnière in the context of functional structuralism, is one of the sources of Fillmore's Case Grammar. It differs from the descriptivist constituency and we prefer its structural trees to those of the minimalism approach, since (along with other advantages) they make it possible without any complications to distinguish between "to the left of" and "above" in the sentence pattern. FGD works with projective trees, which correspond to structures with continuous constituents.
- ¹⁵ Thanks to the substantial progress of computational linguistics (i.e. of the use of both structural and statistical methods in the elaboration of semi-automatic linguistic procedures), a descriptive framework can now be checked as soon as it is implemented and used not only in morphemic ("part-of-speech") corpus tagging, but also in syntactically annotating a large corpus. FGD is being checked now in this sense in the three-level annotation of texts from the Czech National Corpus, see Hajičová (2002).
- ¹⁶ One of the domains in which the concept of markedness has been used as a corner-stone, is the research in the child's language acquisition, see Anderson (2000) and the writings quoted there, especially in analyzing the relationships between parts of speech, naming, reference, syntactic dependency, and predicate-argument structures. However, as long as intonation (prosody) plays a marginal role in child-language research and the predicate-argument relation is not understood as primarily serving the topic-focus articulation of the sentence, the nature of basic linguistic structures and of their role in the child's language acquisition cannot be properly recognized.
- ¹⁷ Note that in the issues of word formation also the tectogrammatical structure of words depends on typological properties.

¹⁸ In a somewhat broader sense naturalness can be also seen in syntax, esp. if issues of economy are taken into account. Thus, it may be understood as suitable for a language system to include much of grammatical information in lexical entries, especially the valency frames.

Bibliographical note: This paper was published in P. Sterkenburg (ed.): *Linguistics Today – Facing a Greater Challenge* (Plenary lectures from the 17th International Congress of Linguists). Amsterdam/Philadelphia: Benjamins 2004, 243–265. The present version of the paper has been slightly enriched.

2.

FREEDOM OF LANGUAGE

Its nature, its sources, and its consequences*

1. Conventions and norms in natural language

1.1. Attitudes towards norms

Natural languages exhibit several layers of norms or conventions, which correspond to D. Lewis' (1969) characterization of 'convention'. However, languages are not rigorously restricted by these norms, most of which lack an explicit formulation (the main exception in this respect concerns point 2 below). The layers of norms differ both in the conditions or factors of their acquisition and in the respective sanctions. Among these layers there are:

1. the **norm** of the language (codified or not), acquired on the basis of innate properties (which perhaps are not as complex as assumed in the Minimalist theory; cf. Sgall 1998; 2000) and of the environment, in part deliberately; the speakers' adherence to this norm is determined first of all by their desire to be well understood without difficulties (without much time and effort to be devoted to decoding and understanding); cases have been found in which the speaker disobeys a rule for the sake of becoming well understood (J. McCawley, p.c.);

2. the **codified** norm of the Standard, consisting in a relatively highly explicit formulation of the basic norm 1, usually modified in a larger or smaller number of points (mostly concerning details of morphemics and similar grammatical issues); the acquisition of the codified norm is based on recognized authorities, including school education, with further prescriptive activities or without them; it should be reflected in linguistics that some speakers adapt more easily and more thoroughly than others

to the requirements of school and of other instances taking care of the Standard (risking, in the extreme, even to become ridiculous in informal situations), while others do not adapt that easily, sticking to the colloquial usage they had adopted in their childhood; as Joos (1961, Section 3) states, it is “the highest compliment possible among mature people” to use intimate style, thus leaving the usual “guardedness”; furthermore, it deserves more attention that conditions for these attitudes differ from one language community to the other (cf. Sgall et al. 1992); the sanctions concern means specific to school (marks or other kinds of evaluation), to editorial activities, criticism, and so on;

3. regularities and conventions in the domain of **discourse**, especially the conventional and conversational implicatures, postulates and maxims of Grice (1975) and others, and the conditions of different kinds of speech acts in the sense of Austin (1962) and Searle (1969; 1983); the usage of these regularities is acquired on the basis of rational attitudes, by imitation, in part also by education; the respective sanctions belong to the domains of easy understanding and of the appropriateness of behavior;

4. in the realm of **style**, attitudes towards 1–3 get differentiated more subtly, which is based on imitation and deliberate learning; here again, relevant criteria concern the appropriateness of behavior, one’s identification with a social group, and the effectiveness of messages addressing members of the same or of another group.

1.2. No strict limitations

An individual’s adherence to any of these degrees of norms and conventions is **not strictly sanctioned**. Languages are sharply restricted neither by their grammatical (lexical, phonological) rules, by a codification of their norm, nor by implicatures or style. Not even **the world** we live in and speak of (or our knowledge of the world) is immediately relevant: not only in fairy tales or when e.g. exaggerating, the speakers are free to depart from the plausibility of the literal content, as far as they believe to be understood.

Further such irregularities concern on the one hand utterances that can occur when discussing language itself, on the other hand when using

conditional, negation, and so on. Thus, questions such as whether wind can open a door, or whether someone might have the right for the sky to be blue, and so on, and so forth, do not actually concern language: this very statement documents that sentences with these and similar combinations of lexical units are not only grammatically well formed, but also display their meaning (in the linguistic sense, which can more precisely be understood as the underlying sentence structure), even though some of such sentences carry only propositions not assigning the value 'true' to any possible world (except in metaphorical uses). Thus (as has been mentioned in our discussions with S. Kuno, Ch. Fillmore, R. Jackendoff and others), e.g., *Martin provided for Mary to live under a blue sky for whole years* cannot express a true statement, but with another modality the possibility of an assignment of the value 'true' may be provided, cf. – with different degrees – such modalities as *he wanted to...*, *he promised...*, *he could not...*

It is true that in several countries patriotic political tendencies have led to the rise of **language laws** concerning protection of the given language against foreign influence and/or limitations concerning the use of its non-standard forms. However, mostly such restrictions are neither far-reaching in their range, nor accompanied by serious sanctions, and in some cases formulations of the laws may appear ridiculous (e.g. when the community councils in a country with large rural areas and strong minorities are formally required to use the standard form of the national language in their sessions). No wonder that such laws are not taken too much seriously by relevant layers of the language communities. It follows that F. de Saussure's comparison of language with the game of chess is not precise, since chess rules differ from language conventions in being compulsory (I owe this remark to J. Peregrin, p.c.).

1.3. Need to be understood

The question then arises what are the main factors that determine the degree of adherence to the norms and conventions of language. Not having the possibility to go into issues such as individual sociolinguistic attitudes or communicative situations, we want to concentrate on one of the factors:

Speakers are free to depart from the norm (or from the plausibility of the literal content), as far as they believe to be **understood** easily enough (cf. Sgall et al. 1986, 32f); the speakers themselves decide which kinds and numbers of figurative expressions, of hyperbolic formulations, of deletions, and so on, are appropriate or tolerable in the given circumstances (with this or that addressee, context, content, etc.).

Among the factors that are immediately relevant in this respect, especially for deletion, there is repetition (including quotation), and also the close connection between utterances representing a direct continuation of preceding context, including coordinated clauses or answers to questions. In utterances displaying the character of such a direct continuation, deletions of many kinds are possible (e.g. *Jim thinks bananas* or *Flowers to Mary* are fully acceptable after *What has Jane bought yesterday?* or in *He gave a bracelet to Nicole; a book to John; flowers to Mary, respectively*). The possibilities of systematic and other deletions are manifold and they are difficult to specify as being restricted by grammatical rules, cf. Chomsky (1975, 119). However, certain language specific restrictions do exist even in the case of such a contextual deletion, cf. the shortened answers in E. *Have you met Harry?* – *I have.* versus Cz. *Potkals Harryho* – *Potkal* [met], where the repetition of either the auxiliary in Czech, or the lexical verb in English (without repeating the object) should be described as grammatically excluded.

Let us add illustrations of these possibilities from different layers of Czech:

- a. Czech is a pro-drop language, i.e. (especially in its relatively formal styles) the subject pronoun has a zero form (in any of the three persons);
- b. as in many languages, also in Czech the whole topic of an utterance in a dialogue may often be left out, cf. (*Kde jsi ho našel?* ‘Where have you found him?’) – *Doma* ‘At home’. (*Proč to tam dáváš?* ‘Why do you put it there?’) – *Aby to nenašel Jirka* ‘For George not to find it’. (*Potkal jsi Milenu?* ‘Have you met Milena’) – *Potkal* (lit. ‘Met’, i.e. ‘I have’); the difference between Czech and English in the last example, which we mentioned above, documents that there are systematic limitations to such deletions in individual languages;
- c. in a non-standard colloquial style there is the option to pronounce individual word forms with most different shortenings; these are limited

neither to unaccented parts of the words, nor to their final or to their beginning parts, as can be documented by the fact that e.g. the word *padesátník* ('fifty-heller coin') can be pronounced either as *pade* or as *sát-ník*, leaving out the other half of the word;

d. moreover, there are the well known cases of minimal deletions, which often represent not only individual events in the process of communication, but penetrate into the language system – cf. the loss of individual phonemes and/or even of marked distinctive features e.g. in the development of the Czech infinitives (*dát* 'to give', rather than *dát'* from *dáti*, etc., for almost all verbs), or the colloquial pronunciation of the name *Botič* (a brook in Prague, pronounced with [t̚] in the standard, but with [t] in most everyday conversations) or of *porád* 'all the time' vs. the standard *pořád*.

The possibilities of deletion (as well as of 'neglected speech' and of other layers of redundancy reduction) are highly manifold and the speakers' choice consists of many options, the repertoire (and hierarchy) of which cannot be easily described systematically. The speakers find their ways by choosing among the different layers of economy and of redundancy. Also the choice of more or less precise expressions belongs here (with underspecification – ambiguity and indistinctness – being possibly resolved by the addressees on the basis of context).

1.4. Tradition and innovations

As Stich (1991) points out, a human being does not have a free choice in deciding which language to adopt. The child is pushed to learn the language of its environment and gets acquainted, step by step, with its basic structure and its idiosyncrasies (cf. also McCawley 1992 on the acquisition of complex syntactic structures). During this process the child has to abandon the systematic, analogical image of the language system it has internalized and to acquire all (or almost all) the historically determined exceptions and idiosyncrasies in the morphemics, syntax, lexicon, phraseology and other layers of the language. Being so equipped, the speaker may act against the tradition of the language and the group uniformity, making use of the principles of freedom, individual independence and creativity. Not only in poetry, in humor,

etc., but also in everyday speech and then in other domains of communication the individual gets aware of his/her individual responsibility both in the choice of means of expression, style, etc., and in creating new ones or new combinations of them.

Speakers often deliberately avoid on the one hand to use superfluous effort in expressing themselves and, on the other, to speak in a way that would not be easily understood and accepted as appropriate (both in what concerns norms of behavior and suitability for the given situation). However, they feel free to use most different kinds of daring innovations and of expressive strengthening (that may give rise to new synonyms or intensify the use of existing ones), to make their speech interesting, inventive, witty, thus maximizing the effectivity of their messages and the attractiveness of their images, i.e. their potential influence.

As the poet writes, “in our language we look for the most elementary freedom – to be able to express our most intimate thoughts. And this is the basis of any freedom” (Seifert 1984). These words certainly concern also the way of expression of our thoughts (and emotions, attitudes, and so on).

2. The sources of the freedom of language

The intrinsic freedom of language, i.e. the freedom of the speaker's choice in all the aspects mentioned (and in others) **follows from the interactive nature** of language. Several aspects of this relationship may be briefly mentioned:

a. speakers are human individuals with their free will and with different backgrounds and psychologies; as Bach (1996) puts it, “humans are not just speaking animals, they are also language-creating animals”; the language norms constitute just a starting point for communication, as Arutjunova (1994) states (in a somewhat different connection); thus, speakers decide more or less consciously and consistently whether to adhere or not to this or that norm in a given point; together with the unlimited **variability** of communicative situations and of (situational and verbal) **contexts** this leads to an indefinite number of possible combinations of sentences occurring in a discourse, which cannot be captured by

any collection of “text patterns” or “textemes” (although various kinds of hierarchies of these combinations certainly can be specified as soon as the factors exhibiting maximal impact are identified); speakers are free also in choosing the degree of their sincerity and forthrightness, as well as in the degree of precision in the semantics of their utterances (cf. esp. Dönninghaus (1999; 2001);

b. already the quite elementary possibility of **repeating** own or others’ formulations, of quoting, recalling, etc., leads to complications in the structure of language; one of the aspects of this effect is the fact that in repetitions and in formulations recalling a part of the preceding co-text the topic-focus articulation of sentences may change, and that any expression may appear in the topic of a sentence; even a focus sensitive particle with its ‘focus’ and scope may then be included in the topic (in the contextually bound part) of an utterance, cf. e.g. (*If everybody expected the weather to deteriorate,*) *then even the most audacious participants might start thinking of finishing the expedition*, where the ‘focus’ of *even* is (expressed by) *the most audacious participants* (which recalls *everybody*), whereas the focus of the sentence is *might start thinking of finishing the expedition* (cf. Hajičová, Partee and Sgall 1998);

c. repetitions may give rise to further aspects of complicating the language structure: they evoke the possibility of deletions and of other **shortenings** mentioned above (as we have seen, there are deletions of most different kinds and layers, from dropping a phoneme down to the ellipsis of the whole topic of a sentence); repetitions are further connected with different kinds of modifications of repeated parts of formulations, since there naturally is a tendency not to repeat literally; here belongs e.g. the already mentioned shortening of answers (not to repeat too much from the question), as well as the expressive strengthening, which supports the use of synonyms;

d. anaphoric and other quotations of (sectors of) co-text allow even for self-reference, certain kinds of which (instantiated by the classical Li-ar’s sentence) represent sentences not expressing any proposition (truth conditions); this is connected with the absence of a clear boundary line between object language and metalanguage (cf. also issues of logophoric contexts, as discussed by A. Kratzer and others).

3. The consequences of the freedom of language

3.1. Variability of expression

Among the consequences of the freedom of language there is, first of all, the well known high degree of **variability** of linguistic forms within language systems, which was discussed already by Mathesius (1911; cf. also the attention he paid in his later writings to what he called the speakers' communicative needs), as one of the corner stones at the beginning of synchronic linguistics. The speaker chooses among the varying items, creates new collocations and words; what starts as an individual usage can find broader response, the individual's choice can lead to a change in the system.

The variability of the expression forms and of their functions leads to the possibility of a '**teleological**' (goal oriented) **development** of the language system. The lexicon steadily gets enriched, with the individual speakers' role coming to the fore, as stressed especially by Ch. Fillmore's (1982; 1985) 'semantics of understanding' (see also Waszakowa 1997) and by cognitive linguistics (which, as Kiefer 1995: 99 points out, opens new horizons, although it should not be understood as substituting older trends in linguistic research). The lexicon keeps growing in always new idiomatic collocations and complex predicates, technical terms and other neologisms, which come from different sources; some of these are:

1. the historically emerging need to have a denomination for a new concept; this concerns technical terms, but also other words, be they borrowed from other languages (often in a narrower sense than they have in the source language, cf. Cz. *kolaborant*, which does not refer to a collaborator in general, but to a quisling) or created using productive means of word formation;

2. an emotionally marked synonym is first used individually for the sake of marking one's speech as specific, then it gets stabilized in the language system as bearing a stylistic value of expressiveness, and eventually it may substitute the original, stylistically neutral word, cf. examples such as French *tête* 'head' from Lat. *testa*, which already in Colloquial ("Vulgar") Latin substituted the older *caput*, or Cz. *strašný, ohromný* in the meaning of 'big'. It is important to see that also grammatical means

steadily get enriched both in what concerns new forms of expression and their functions: new complex prepositions and conjunctions emerge (i.e. get grammaticalized), etc.

It is important to see also the emergence of new complex predicates; in contemporary Czech this process appears to be quite intensive, cf. such recently spreading verb complexes as e.g. in *To jsme se neměli v úmyslu učit*, lit. 'This we had-not-the-intention-to learn' (built in analogy to *To jsme se nechtěli učit* 'This we did not want to learn'), *Dlužili jsme nějaké daně židovské obci* 'We owed some taxes to the Jewish Community' (Ota Pavel), *Jen nám radili studovat chemii* 'They only recommended us to study chemistry' (cf. Hajičová et al. 1998 on 'Taglicht's sentences'). Certain properties of these and other types of collocations, connected with the word order positions of their parts, document their lexicalized character; their parts are connected with each other more closely than just by regular syntactic relations, which means that they can best be understood as idiomatic phraseological units.

3.2. Language development

Another layer of individual decisions determines, in many languages (i.e., in certain stages of their slow development based on choice in variation), the repeated reduction of grammatical morphemes, which is connected with the emergence of new function words, cf. the typological 'Spirallauf' of G. von der Gabelentz. This typological spiral does not cover the whole domain of typological change and cannot be understood as the backbone of language development as a whole (for which the just mentioned enrichment in most different layers of language constitutes a more important line). In any case, the typological spiral plays a significant role, and it can be seen that this aspect of development is closely connected both with the speakers' tendency for economy and the necessity for them to express themselves with a certain degree of redundancy to be easily understood.

The teleological aspect of language development, and thus also of language structure, was highly stressed by R. Jakobson, N. Trubetzkoy and other members of the classical Prague School; cf. esp. Trost (1989), and also Leška (1987;1994), who discusses sources of the looseness of the

system of language in this connection; as Sériot (1999a;b) points out, this view was rooted in the influence of older Russian traditions. Although this aspect of language and its development is not strong enough to enable the linguists to predict individual changes or to specify the respective regulating mechanisms, it clearly is present, and, as we have just mentioned, it is relevant both for the synchronic study of language and for a description of its development.

3.3. Universality and its limitations

The free variability of language means and their functions is a precondition for its **universality**. Without a certain degree of indistinctness of language meanings (i.e. of the units of the layer of functions of expressions in the language system) it would not be possible to capture with limited means the unlimited range of the world we perceive and speak of, cf. Marty (1908), Karcevskij (1929), Mathesius (1942) or Putnam's (1975) 'stereotypes' and the Gricean (1975; 1982) notions of 'licence' and 'ideal limits'. The fuzzy or indistinct demarcation of linguistic meanings is not only a precondition of the universality of natural language, but also one of its consequences: the complexity of language makes it impossible for any exact rules to be precisely applied in communication.

The universality of natural language is, in fact, restricted, cf. especially Keenan's (1975) findings on relative clauses: neither in standard Czech nor in English a sentence with a relative pronoun renders exactly what colloquial Czech expresses by the particle cooccurring with an anaphoric pronoun e.g. in *Ten pán, co jsem jeho a jeho dceru včera potkal*, lit. 'The gentleman what him and his daughter I yesterday met' (Czech shares this construction with certain south-German dialects and e.g. with Hebrew). The source of another layer of semantic limitations is constituted by the systemic ordering of complementations, which always is present in the 'underlying word order' in the focus of a sentence (cf. Sgall et al. 1986); e.g. the Czech sentence *Pavel vždycky působil svými výroky nějaké obtíže* is translated precisely neither by *Paul always caused some DIFFICULTIES with his statements* (where the *with*-group belongs to the topic, although in Cz. it clearly is a part of focus), nor by *Paul always caused some difficul-*

ties with his statements (with normal intonation, i.e. with the intonation center in the final position, where the difference in the degrees of communicative dynamism – and thus also in the preferred scopes – of the two parts of the focus gets inversed). Due to the difference in the systemic orderings of the two languages an actually true translation is impossible in such cases.

Still another layer of limitations of the universality of natural language can be found in the fact that speakers may find lacunes in the repertoires of lexical units when they want to characterize certain objects or situations. Thus, e.g., we do recognize individual faces and distinguish them from each other, but we do not always find language means to describe the individual differences in detail (I owe this observation to B. H. Par-tee, p.c.).

As analytic philosophy tells us (see e.g. Peregrin 1992), we are limited by the structure of our language in perceiving and understanding the world. To quote a poet again, we may recall Frynta (1993): “A nadto je nám řeč ... metodou orientace ve veškerém jsoucnu” (‘And, moreover, language is for us a method of orientation in the whole universe’). Thus the limits of the freedom of language seem not to be narrower than those of the freedom of human mind.

3.4. Complexity of language

Even with such limitations, the freedom of language together with its relative universality leads to its extreme **complexness**. It is not easy to imagine a system that would be fundamentally more complex than natural language. Let us just recall that this does not exempt the linguist from his obligation to describe language as being (seemingly paradoxically) maximally complex and, at the same time, organized along a core simple enough for a child to be acquired without explicit learning. We are committed to look for as simple a patterning of this core as possible, to avoid the assumption of an unnecessarily complex collection of “innate ideas”; cf. Sect. 4 below for a discussion of this issue.

With its freedom, language gets into situations which were not present in the primary conditions of its existence and for which its structure cannot easily be adapted, i.e. into situations causing **collisions** on most

different levels. Many of such collisions come into being with new requirements yielded by the development of civilization, be it in its technical aspects or in requirements of different layers of communication in the society.

This concerns individual lexical units, cf. the problems connected with the English pronouns *he* and *she*, or with *you* (with which the Plural function often has to be made clear by the use of *you all*, *you guys*, etc.). In German, similar difficulties accompany the Plural forms of the 'polite' pronoun *Sie* vs. the Plural *Ihr*; cf. also noun forms such as German *LehrerInnen* or Czech *učitelé/ky*, as compared to E. *teachers*: the E. nouns of this group do not distinguish gender, so that after an occurrence of e.g. *his neighbor* it may be difficult to decide which gender of the coreferring pronoun should be used (and whether we face an ambiguous noun).¹

Grammar rules often show collisions in cases of morphemic exceptions, which with frequent words are broadly recognized and used (e.g. the suppletion of Cz. *člověk* – *lidé* '(hu)man – people' or Loc.Sing. *ve dne* vs. *o dni/dnu* 'day' (with different prepositions), but may recede if they concern relatively rarely occurring words, so that many speakers are not aware of their existence.

In Czech, similar collisions concern the declension of proper names belonging to restricted inflectional classes; some of the family names originally were identical to common nouns and now differ from them in their oblique cases, such as *Hraběho* vs. *hraběte* (Gen./Acc. of 'earl', *Švece* vs. *ševce* (Gen./Acc. of 'shoemaker'); other cases are such as e.g. the Voc. of the local name *Hradec*, which according to the respective inflectional paradigm would be **Hradče*; however, due to the non-productivity of this paradigm (at least with inanimate nouns), the actually used (although of course not frequent) form is *Hradci*, analogical to the more productive declension class of *stroj* 'machine'. Another group with similar collisions concerns the feminine family names of foreign origin, which do not always display the regular Cz. suffix *-ová*: with complex names such as *Thurn-Taxisová* the suffix usually is used only at the latter part of the name (although according to the codified norm it should occur also with its first part); with some names ending in a vowel, Czech speakers are used to forms such as *Dijkstra* (although e.g. *Parteeová* does occur, perhaps due to the fact that it is used first of all by linguists).²

It is well known that along with these and similar extreme cases, there are many other kinds of collisions between different regularities in language structure. The influence of other languages may cause collisions such as those concerning the declension forms of different Czech loan nouns, or pronunciation and spelling in English, and so on. Other collisions concern conversion and classification of word classes, as well as their syntactic properties and conditions of their occurrence; cf. e.g. Cz. constructions such as *já rád sladké* 'I like sweet (things to eat)', with a word that is morphemically characterized as pronoun or adjective, but occupies the syntactic position of a finite verb (even with the corresponding analytic preterite *já jsem rád*).

The hierarchy of the levels of language gets complicated by collisions of many different kinds. Let us just recall a few examples that are more or less typical for many languages:

a. There are cases in which an opposition that basically pertains to morphemics is used more or less systematically as also relevant for meaning. Although function words basically occupy the same position as function morphemes (their analytic character belongs to the level of morphemics, rather than to the underlying structure), their use occasionally may obtain semantic relevance. Thus, it deserves further discussion whether (or, under which conditions) in certain kinds of coordination structures a preposition perhaps is repeated only if relevant semantic conditions are met; this may concern the Czech construction *pro předsedu a pro ředitele* 'for the chairman and for the director', which is used only as referring to two different persons (K. Oliva, p.c.).

b. In the so-called 'second instance' sentences (or, perhaps, in all sentences with a specific 'narrow focus'), the intonation center (sentential stress) may fall on a syllable that prototypically does not carry the word stress (on a prefix in English, on an ending in Czech, etc.). Such a marked accentuation indicates that only one of the derivational or grammatical values carried by the given word form is in the focus, all the rest of the sentence belonging to its topic, referring to 'given' information, cf. Cz. (*Ne že to udělá.*) *On to udělAL* '(It is not only that he will make it.) He HAS made it'.

c. Direct speech often is described as an object clause dependent on the introducing verb. While this approach is adequate for sentences