

Pragmatic Organization of Discourse
in the Languages of Europe



Empirical Approaches to Language Typology



EUROTYP 20-8

Editors

Georg Bossong
Bernard Comrie

Mouton de Gruyter
Berlin · New York

Pragmatic Organization of Discourse in the Languages of Europe

edited by

Giuliano Bernini

Marcia L. Schwartz

Mouton de Gruyter
Berlin · New York

Mouton de Gruyter (formerly Mouton, The Hague)
is a Division of Walter de Gruyter GmbH & Co. KG, Berlin.

© Printed on acid-free paper which falls within the guidelines of the
ANSI to ensure permanence and durability.

Library of Congress Cataloging-in-Publication Data

Pragmatic organization of discourse in the languages of Europe / edited
by Giuliano Bernini, Marcia L. Schwartz.

p. cm. — (Empirical approaches to language typology. EURO TYP ;
20-8)

Includes bibliographical references and index.

ISBN-13: 978-3-11-015753-6 (hardcover : alk. paper)

ISBN-10: 3-11-015753-5 (hardcover : alk. paper)

1. Europe — Languages. 2. Pragmatics. 3. Grammar, Comparative
and general — Syntax. 4. Discourse analysis. I. Bernini,
Giuliano. II. Schwartz, Marcia L. III. Series.

P380.P73 2006

306.44094—dc22

2005031059

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche
Nationalbibliografie; detailed bibliographic data is available in the
Internet at <<http://dnb.ddb.de>>.

ISBN-13: 978-3-11-015753-6

ISBN-10: 3-11-015753-5

© Copyright 2006 by Walter de Gruyter GmbH & Co. KG, D-10785 Berlin.

All rights reserved, including those of translation into foreign languages. No part of this book
may be reproduced or transmitted in any form or by any means, electronic or mechanical,
including photocopy, recording or any information storage and retrieval system, without per-
mission in writing from the publisher.

Typesetting: Christoph Eyrich, Berlin.

Printing and binding: AZ Druck und Datentechnik GmbH, Kempten (Allgäu).

Printed in Germany.

The present volume is one of a series of nine volumes in which the results of the European research project “Typology of Languages in Europe” (EUROTYP) are published. The initiative for a European project on language typology came from a proposal jointly submitted to the European Science Foundation (ESF) by Johannes Bechert (University of Bremen), Claude Buridant (University of Strasbourg), Martin Harris (University of Salford, now University of Manchester) and Paolo Ramat (University of Pavia).

On the basis of this proposal and following consultations with six experts the Standing Committee for the Humanities of the ESF decided to organize a workshop (Rome, January 1988), in which this idea was further explored and developed. The results of this workshop (published by Mouton, 1990) were sufficiently encouraging for the Standing Committee to appoint a preparatory committee and entrust it with the tasks of drawing up a preliminary proposal, of securing interest and participation from a sufficiently large number of scholars and of finding a suitable programme director. The project proposal formulated and sent out by Simon Dik (University of Amsterdam) as chair of this committee met with very supportive and enthusiastic reactions, so that the Standing Committee for the Humanities recommended the funding of a planning stage and the General Assembly of the ESF approved a year zero (1989) for an ESF Programme in Language Typology.

During this planning phase all major decisions concerning the management structure and the organisation of the work were taken, i.e., the selection of a programme director, the selection of nine focal areas around which the research was to be organized, the selection of a theme coordinator for each theme and the selection of the advisory committee.

The first task of the programme director was to draw up a definitive project proposal, which was supplemented with individual proposals for each theme formulated by the theme coordinators, and this new proposal became the basis of a decision by the ESF to fund the Programme for a period of five years (1990–1994).

Language typology is the study of regularities, patterns and limits in crosslinguistic variation. The major goal of EUROTYP was to study the patterns and limits of variation in nine focal areas: pragmatic organization of discourse, constituent order, subordination and complementation, adverbial constructions, tense and aspect, noun phrase structure, clitics and word prosodic systems in the languages of Europe. The decision to restrict the investigation to the languages of Europe was imposed for purely practical and pragmatic reasons. In the course of the project an attempt was made, however, to make as much sense of this restriction as possible, by characterizing the specific features of European languages against the background of

non-European languages and by identifying areal phenomena (*Sprachbünde*) within Europe.

More specifically, the goals of the EUROTYP project included the following:

- to contribute to the analysis of the nine domains singled out as focal areas, to assess patterns and limits of cross-linguistic variation and to offer explanations of the patterns observed.
- to bring linguists from various European countries and from different schools or traditions of linguistics together within a major international project on language typology and in doing so create a new basis for future cooperative ventures within the field of linguistics. More than 100 linguists from more than 20 European countries and the United States participated in the project.
- to promote the field of language typology inside and outside of Europe. More specifically, an attempt was made to subject to typological analysis a large number of new aspects and domains of language which were uncharted territory before.
- to provide new insights into the specific properties of European languages and thus contribute to the characterization of Europe as a linguistic area (*Sprachbund*).
- to make a contribution to the methodology and the theoretical foundations of typology by developing new forms of cooperation and by assessing the role of inductive generalization and the role of theory construction in language typology. We had a further, more ambitious goal, namely to make a contribution to linguistic theory by uncovering major patterns of variation across an important subset of languages, by providing a large testing ground for theoretical controversies and by further developing certain theories in connection with a variety of languages.

The results of our work are documented in the nine final volumes:

Pragmatic Organization of Discourse in the Languages of Europe

(edited by G. Bernini)

Constituent Order in the Languages of Europe (edited by A. Siewierska)

Subordination and Complementation in the Languages of Europe

(edited by N. Vincent)

Actance et Valence dans les langues d l'Europe (edited by J. Feuillet)

Adverbial Constructions in the Languages of Europe (edited by J. van der Auwera)

Tense and Aspect in the Languages of Europe (edited by Ö. Dahl)

Noun Phrase Structure in the Languages of Europe (edited by F. Plank)

Clitics in the Languages of Europe (edited by H. van Riemsdijk)

Word Prosodic Systems in the Languages of Europe (edited by H. van der Hulst)

In addition, the EUROTYP Project led to a large number of related activities and publications, too numerous to be listed here.

At the end of this preface, I would like to express my profound appreciation to all organizations and individuals who made this project possible. First and foremost, I must mention the European Science Foundation, who funded and supported the

Programme. More specifically, I would like to express my appreciation to Christoph Mühlberg, Max Sparreboom and Geneviève Schauinger for their constant and efficient support, without which we would not have been able to concentrate on our work. I would, furthermore, like to thank my colleague and former assistant, Martin Haspelmath, and indeed all the participants in the Programme for their dedication and hard work. I finally acknowledge with gratitude the crucial role played by Johannes Bechert and Simon Dik in getting this project off the ground. Their illness and untimely deaths deprived us all of two of the project's major instigators.

Ekkehard König, Programme Director

Contents

General Preface	v
---------------------------	---

Contributors	xxiii
------------------------	-------

I. General issue

<i>Giuliano Bernini</i>	
Introduction	3

<i>Jim Miller and M. M. Jocelyne Fernandez-Vest</i>	
Spoken and written language	9

II. From function to form

<i>Elena Maslova and Giuliano Bernini</i>	
Sentence topics in the languages of Europe and beyond	67

<i>Jim Miller</i>	
Focus in the languages of Europe	121

<i>Yaron Matras and A. Machtelt Bolkestein</i>	
Deixis and anaphora: Some case studies	215

<i>Hans-Jürgen Sasse</i>	
Theticity	255

III. From form to function

<i>Alan Cruttenden</i>	
The de-accenting of given information: A cognitive universal?	311

<i>Rosanna Sornicola</i>	
Interaction of syntactic and pragmatic factors on basic word order in the languages of Europe	357

<i>Christoph Schroeder</i>	
Articles and article systems in some areas of Europe	545

Indexes

Subject index 615

Language index 619

Author index 625

Contents detailed

General Preface	v
Contributors	xxiii

I. General issue

Giuliano Bernini

Introduction	3
------------------------	---

Jim Miller and M. M. Jocelyne Fernandez-Vest

Spoken and written language	9
1. Introduction	9
2. Distinguishing spoken and written language	9
2.1. Genres	9
2.1.1. Biber's textual dimensions	10
2.1.2. Spontaneous spoken language, written language, and education	11
2.2. Previous work on spoken and written language	13
2.2.1. General grammatical properties of spontaneous spoken language	13
2.2.2. General discourse properties of spontaneous spoken language	14
2.2.3. Analyses contra a special status for spontaneous spoken language	15
3. Sentences and clauses	16
3.1. Introduction	16
3.2. System sentences and text sentences	16
3.3. Sentences in spoken texts	19
3.3.1. Introduction	19
3.3.2. Intuitions about sentences in spoken language	20
3.3.3. The sentence: A changing concept	20
3.4. Sentences and syntactic analysis	24
3.4.1. The clause as the central unit of syntactic analysis	24
3.4.2. Dependencies crossing sentence boundaries	27
3.4.3. Sentences, clause, and distribution	27
4. "Syntactic fragments," competence, and performance	29

5. Clause and clause complex	33
6. Noun phrases	35
7. Discourse	36
8. Typology	44
8.1. Introduction	44
8.2. Relative clauses, spoken and written language, and typology . . .	46
8.2.1. Typological overview	46
8.2.2. The order of the head NP and the modifying relative clause	46
8.2.3. How the head noun is reflected in the relative clause . .	48
8.2.4. English relative clauses	49
8.2.5. Relative clauses in Russian	54
9. Conclusion	56
Notes	56
References	61

II. From function to form

Elena Maslova and Giuliano Bernini

Sentence topics in the languages of Europe and beyond	67
1. Introduction	67
2. The hypothesis of existence for sentence topic: Pro and contra	68
3. Phenomenology of sentence topics: An overview	71
3.1. Packaging topics	71
3.2. Hanging topics	73
3.3. Topic constructions	76
3.4. Non-nominal topics	80
4. Semantic grounding of the topic-comment structure	83
4.1. Introductory remarks	83
4.2. Primary participants	84
4.3. Reference points	88
4.4. Spatial frames	93
4.5. Three semantic templates vs. three aspects of topicality	96
5. Language-internal and crosslinguistic variation in topic encoding . . .	100
5.1. Two dimensions of topic prominence	100
5.2. Competition between topic-encoding templates	103
5.3. Interaction between topic-encoding templates	108
5.4. Conclusion: Universal constraints on topic encoding	113
Notes	114

Jim Miller

Focus in the languages of Europe	121
1. Introduction	121
2. Theory and data	122
2.1. Basic concepts	122
2.2. Data collection	123
3. Results	127
4. Focus	129
4.1. Overview	129
4.2. Halliday	130
4.3. Chafe: Components of contrastiveness	130
4.4. Dik: Extended notion of focus	131
4.5. Vallduví	132
4.6. Grosz and Sidner: Focus spaces and transition	133
5. New individuals and sets	134
5.1. Focus spaces, file cards and entities	134
5.2. English	135
5.2.1. New entities in written texts	135
5.2.2. NP-clause	135
5.2.3. SEE and KNOW	136
5.3. German	137
5.4. French	138
5.5. Russian	139
5.6. Hungarian	142
5.7. Finnish	142
5.8. Turkish	142
5.9. Irish and Scottish Gaelic	142
5.10. Bulgarian	143
5.11. Topic-prominent and subject-prominent languages	144
6. New propositions	145
6.1. English	145
6.1.1. <i>the thing is</i>	145
6.1.2. SEE introducing propositions	146
6.2. German	147
6.3. Greek	148
6.4. Russian	148
6.5. French	151
7. Russian word order	151
7.1. Introduction: Siewierska and Bivon	151
7.2. Word order in spoken and written Russian and Polish	152
7.3. Clause position in written Russian	154

7.4. Word order as a highlighting device	156
8. Reduction in salience	157
8.1. Overview	157
8.2. Ellipsis in Russian	158
8.2.1. Null subject NPs	158
8.3. Null object NPs	160
8.3.1. Null NPs in spoken and written Russian	162
8.4. Salience-reducing constructions	162
9. Highlighting devices (nonpresentational)	165
9.1. Discourse particles	165
9.2. Word order	166
9.3. Contrastive focus position	166
9.4. Syntax and intonation	170
9.5. Clefts	171
9.5.1. Definition of cleft constructions	171
9.5.2. Languages without clefts	174
9.5.3. Languages with clefts	179
10. Spontaneous map task dialogues	188
10.1. Introduction	188
10.2. English	189
10.3. German	190
10.4. Catalan	193
10.5. Hungarian	198
10.6. Bulgarian	200
11. Conclusions	203
Notes	205

Yaron Matras and A. Machtelt Bolkestein

Deixis and anaphora: Some case studies	215
1. Introduction	215
2. Extratextual deixis, intratextual deixis and anaphora	216
3. Extratextual deixis, intratextual deixis, and anaphora in Classical Latin	220
3.1. Extratextual deixis	220
3.2. The system of intratextual deixis and anaphora in Latin	221
4. Situational deixis, intratextual deixis, and anaphora in modern Dutch	230
4.1. Situational deixis in Dutch	230
4.2. Intratextual deixis and anaphora in Dutch	231
5. Deixis and anaphora in Romani and Modern Hebrew	239
5.1. Romani	239
5.2. Hebrew	246

6. Concluding remarks	249
Notes	250

Hans-Jürgen Sasse

Theticity	255
1. Introduction	255
1.1. The phenomenon	255
1.2. Previous research	257
1.3. The present research strategy	262
2. Constructions	264
2.1. Formal types of “special” constructions in the domain of theticity	264
2.2. Distribution of formal types across languages	267
2.3. Text frequency of “special” constructions	269
2.4. Polysemy of constructions	271
2.4.1. VS constructions	271
2.4.1.1. Narrow focus	271
2.4.1.2. Inversion	272
2.4.2. SAcc and Split constructions	273
2.5. Language-specific restrictions	274
2.5.1. Global syntactic patterns	274
2.5.2. Monoargumentality	277
2.5.3. Definiteness and animacy	279
2.5.4. Local restrictions	280
3. Theticity and discourse functions	280
3.1. The relevant discourse-functional domains	280
3.1.1. The annuntiative function	281
3.1.2. The introductive function	284
3.1.3. The interruptive function	285
3.1.4. The descriptive function	286
3.1.5. The explanative function	287
3.2. Similarities and differences among the discourse functions	288
3.3. Coverage of functions by construction types in the different languages	290
4. The “connective” complex	294
5. The impact of lexical semantics	297
6. Summary, conclusions, and perspectives for future research	299
Notes	302

III. From form to function*Alan Cruttenden*

The de-accenting of given information: A cognitive universal?	311
1. Introduction	311
2. De-accenting and given information	314
3. De-accenting and re-accenting	315
4. Settings and responses	318
4.1. Setting-response 1	318
4.2. Setting-response 2	320
4.3. Setting-response 3	320
4.4. Setting-response 4	321
4.5. Setting-response 5	322
4.6. Setting-response 6	323
4.7. Setting-response 7	324
4.8. Setting-response 8	324
4.9. Setting-response 9	325
4.10. Setting-response 10	326
5. Conclusions	327
5.1. Re-accenting in Romance and other languages	327
5.2. Contrast	329
5.3. Discourse and grammar	331
A Albanian	333
B Arabic (Tunisian)	335
C French	337
D German	339
E Greek	340
F Italian	342
G Lithuanian	344
H Macedonian	345
I Russian	347
J Spanish	349
K Swedish	351
Notes	352

Rosanna Sornicola

Interaction of syntactic and pragmatic factors on basic word order in the languages of Europe	357
1. Theoretical and methodological foundations	357
1.1. A pragmatic study of word order	357
1.2. Word order between competence and performance	358

1.3.	Microscopy and macroscopy	361
1.4.	The languages of the corpus	361
1.5.	The method of gathering data	363
1.6.	Topology and syntax: Towards a geometry of the sentence	364
1.6.1.	Notions of order	364
1.6.2.	Models of order as a result of movement	365
1.6.3.	The concept of position	366
1.6.4.	Representation by fields	368
1.7.	The descriptive traditions	368
1.7.1.	Diversity in traditions of word order study	368
1.7.2.	The d-configurationality model	370
1.7.2.1.	Heltoft's criticism of generative models	370
1.7.2.2.	Configurationality and d-configurationality	372
1.8.	Form and function in the study of word order	374
1.8.1.	Grammatical functions, semantic functions, pragmatic functions	375
1.8.1.1.	The syntactic and prosodic realization of pragmatic functions	377
2.	Word order in sentences with two arguments: The relationship between position, syntactic function, and pragmatic function	379
2.1.	TOPIC position	379
2.2.	Languages with O in preverbal position in the basic order and FOCUS in preverbal position	379
2.2.1.	Basque and Turkish	381
2.3.	Problematic cases concerning the relationship between position, grammatical function, and pragmatic function	384
2.3.1.	Georgian	384
2.3.2.	Hungarian	387
2.4.	The interplay of position, grammatical functions, and pragmatic functions in languages with O in postverbal position	389
2.4.1.	General considerations about the position of the FOCUS constituent in SVO languages	389
2.4.2.	SVO languages	391
2.4.3.	SVO languages with highly flexible word order	393
2.4.3.1.	Estonian and Finnish	393
2.4.3.2.	The Slavonic languages	397
2.4.4.	SVO languages with weak word order flexibility	403
2.4.4.1.	The Romance languages	403
2.4.4.2.	English	409
2.4.4.3.	Some remarks on Italian and Spanish	411
2.4.5.	V-2 languages	413

2.4.5.1.	The role of scientific traditions in the formation of the V-2 concept	413
2.4.5.2.	The Germanic languages	414
2.4.5.3.	Vorfeld, Mittelfeld, Nachfeld	416
2.4.5.4.	Some non-V-2 structures	417
2.4.5.5.	German	419
2.4.6.	V-initial languages	429
2.4.6.1.	Irish and Welsh	429
2.4.6.2.	Some problems concerning the relationship between position, grammatical function, and pragmatic function	429
2.4.6.3.	Concurrent SVO types	431
2.4.6.4.	Properties of position P_1	432
2.4.6.5.	More on position P_1 : A controversial type with NP in P_1	434
2.4.6.6.	A few concluding remarks	437
3.	Word order in sentences with one-argument verbs	442
3.1.	Monoargumentality and word order patterns	442
3.1.1.	Constructions with one-argument verbs	442
3.1.2.	Between the microscopic and the macroscopic	445
3.2.	SVO languages	447
3.2.1.	Indo-European languages	447
3.2.1.1.	The (X)VS pattern	449
3.2.2.	The Balto-Finnic languages	452
3.3.	VSO languages	452
3.4.	SOV languages	453
3.5.	Hungarian	455
3.6.	Between universal tendencies and typological conditions	456
3.6.1.	Structural factors and statistical regularities in unstable orders	456
3.6.2.	Some problems with generative models	457
3.6.3.	Typological conditions	460
4.	V or sentence in FOCUS	461
4.1.	V in FOCUS	461
4.2.	Word order when sentence is all-in-FOCUS	464
4.2.1.	A few conclusions on typological properties of all-in-FOCUS structures	466
5.	The influence of semantic or textual features on word order	466
5.1.	The influence of the semantic features animacy, definiteness, and referentiality on word order	466
5.2.	Textual relations	470
5.2.1.	The effect of the features GIVEN and NEW on word order	470

5.2.2.	The distribution of GIVEN and NEW in SVO languages	471
5.2.3.	The distribution of GIVEN and NEW in VSO languages	474
5.2.4.	The distribution of GIVEN and NEW in SOV languages	475
6.	Conclusions	476
6.1.	Universal tendencies and the typology of word order	476
6.1.1.	The “written” vs. “spoken” parameter	476
6.1.2.	The historical conditions	477
6.1.3.	Structural conditions	478
6.1.4.	Criteria for determination of the sentence space	479
6.1.5.	Some principles of word order organization	481
6.2.	Word order and the asymmetry of pragmatic functions	486
6.2.1.	The relationship between position, grammatical function, and pragmatic function	487
6.2.2.	Codification of FOCUS by means of word order	489
6.2.3.	The relationship between the grammatical function O and FOCUS	491
6.2.4.	Some possible typological correlations with the parameter “FOCUS in P_k ”	492
6.2.4.1.	Hypothesis 1	492
6.2.4.2.	Hypothesis 2	495
6.2.4.3.	A third hypothesis	496
Notes		497

Christoph Schroeder

Articles and article systems in some areas of Europe	545
1. Introduction	545
2. The categories	546
2.1. Definiteness	546
2.1.1. Identifiability	546
2.1.2. Identifiability based on anaphora	547
2.1.3. Relationship of identity	548
2.1.4. Identifiability and situational context	549
2.1.5. Generic NPs and proper names	550
2.2. Indefiniteness, specificity, referentiality, nonreferentiality	551
3. Articles	552
3.1. Means of expression of (in)definiteness	552
3.2. Definite articles	553
3.3. Demonstratives and possessives as articles	553
3.4. Syntacticization trends of definite articles	555
3.5. Indefinite articles	555
3.6. From numeral to indefinite article	556

3.7.	Emergent indefinite article	557
3.8.	Pragmatic indefinite article	557
3.9.	Referential indefinite article	558
3.10.	Extended indefinite article	559
4.	The survey	560
4.1.	The north: Two definite articles	560
4.1.1.	North Frisian and Low German	560
4.1.2.	Danish dialects of Jutland	562
4.1.3.	Standard Danish	562
4.1.4.	Swedish, Norwegian, Faroese	564
4.2.	Excursus 1: Two definite articles in the Low German dialects of the North Sea region	565
4.3.	Transitions I	570
4.3.1.	Adjectival endings	570
4.3.2.	Attributive and predicative forms of adjectives	571
4.3.3.	Baltic languages	573
4.3.4.	Balto-Finnic languages	575
4.3.5.	Slovene, Serbian, Croatian	576
4.4.	The Balkan Sprachbund and the south of it	577
4.4.1.	Macedonian	577
4.4.2.	Bulgarian	578
4.4.3.	Albanian, Rumanian	579
4.4.4.	Greek	582
4.4.5.	Article reiteration in Greek and Maltese	583
4.4.6.	Indefinite articles	584
4.5.	Offshoots	584
4.6.	The eastern margins of Europe: Possessive suffixes	585
4.6.1.	Uralic	585
4.6.2.	Komi	586
4.6.3.	2nd person and 3rd person possessive suffixes	587
4.6.4.	Possessive suffixes as emerging articles	588
4.6.5.	Indefinite articles in Uralic	591
4.6.6.	Turkic	591
4.6.7.	Adverbs and adjectives of place and time	592
4.6.8.	Indefinite articles in Turkic	592
4.7.	Excursus 2: The possessive suffix 3 SG in Turkish – a definite article?	594
4.7.1.	Agreement	594
4.7.2.	Relational possessive	595
4.7.3.	Compound marker	596
4.7.4.	Domains of use of possessive suffixes	596

4.7.5. Categorical status of possessive suffixes	597
5. Conclusions	598
5.1. Functions of definite articles	598
5.2. Continua of development of articles	599
5.2.1. Continuum 1: Indefinite articles	600
5.2.2. Continuum 2: Definite articles	600
5.2.3. Continuum 3: Possessive suffixes	601
Notes	601

Indexes

Subject index	615
Language index	619
Author index	625

Contributors

Giuliano Bernini
Università di Bergamo
Facoltà di Lingue e letterature straniere
Piazza Rosate 2
24129 Bergamo
Italy
E-mail: giuliano.bernini@unibg.it

A. Machtelt Bolkestein
passed away on October 21st, 2001,
in Amsterdam, The Netherlands.

Alan Cruttenden
Professor emeritus of Phonetics, University of Manchester
Fellow of the Phonetics Laboratory, University of Oxford
Postal Mail:
31 Picklers Hill
Abingdon
Oxon OX14 2BB
Great Britain
E-mail: cruttenden@yassoo.demon.co.uk

M. M. Jocelyne Fernandez-Vest
Linguistique Générale et Ouralienne
E. P. H. E. – Sorbonne
Esc. E, 1er étage
45-47 rue des Ecoles
75005 Paris
France
E-mail: mmjfv@ext.jussieu.fr

Yaron Matras
School of Languages, Linguistics and Cultures
University of Manchester
Oxford Road
Manchester M13 9PL
Great Britain
E-mail: yaron.matras@manchester.ac.uk

Elena Maslova
2000 Walnut Ave, J307
Fremont, CA 94538
USA
E-mail: Maslova@jps.net

Jim Miller
Department of Applied Language Studies and Linguistics
University of Auckland
Private Bag 92019
Auckland
New Zealand
E-mail: je.miller@auckland.ac.nz

Hans-Jürgen Sasse
Institut für Linguistik
Abt. Allgemeine Sprachwissenschaft
Universität zu Köln
50923 Köln
Germany
E-mail: hj.sasse@uni-koeln.de

Christoph Schroeder
Istanbul Bilgi University
Faculty of Science and Letters
Kuştepe Campus
İnönü Cad. 28
34387 Şişli/ Istanbul
Turkey
E-mail: Schroeder@bilgi.edu.tr

Rosanna Sornicola
Dipartimento di Filologia Moderna
Universita' di Napoli Federico II
Via Porta di Massa 1
80133 Napoli
Italy
E-mail: sornicol@unina.it

I. General issue

Introduction

The eight contributions comprised in this volume of the EUROTYP series are devoted to the investigation of single areas in the field of the pragmatic organization of discourse in the languages of Europe. The contributions are grouped into three main parts, according to the issues addressed (general vs. particular) and to the approach adopted with respect to the dichotomy of function and form.

In Chapter 1 (“Spoken and written language”) of Part I, which is devoted to “General issues,” Jim Miller and Jocelyne M. M. Fernandez-Vest investigate the different organization patterns of discourse resulting from the spoken vs. written medium of the transmission of verbal messages. The diamesic dimension is actually a continuum where different text types or genres distribute between the extreme points of *impromptu* speech on the “spoken” side of the continuum and high literature on the “written” side of the continuum in a non-clear-cut, but rather gradient way. The different factors conditioning language production along the diamesic dimension have consequences for the theoretical status of traditional units of analysis and for typology. As for traditional units of analysis, sentence is shown to be a low-level discourse unit mainly of written language; along the spoken side of the diamesic dimension and, most of all, at the endpoint of spontaneous spoken language, clause appears to be a more relevant unit of discourse analysis along with phrase, allowing a better understanding of the (apparent) fragmentary structure of spoken discourse. As for typology, different syntactic patterns appear to characterize the expression of the same function in spoken and written language, as in the case of conditional and relative clauses, among others. Occurrence in different languages and persistence over time of the distribution of different constructions for the same function makes the diamesic dimension a relevant parameter in the analysis of discourse organization in a typological approach.

Part II (“From function to form”) comprises four contributions investigating in a problematic perspective some of the functions traditionally considered in studies of the pragmatic organization of discourse and the range of expressions found for them in European languages.

The first function considered is that of topic, dealt with by Elena S. Maslova and Giuliano Bernini (“Sentence topics in the languages of Europe and beyond”). The contributors account in a unified fashion for the internal and crosslinguistic variation found for topic expressions in different languages and often thought to challenge the

validity of a relatively vague notion as that of topic as “what the sentence is about.” The topic status of a referent is claimed to be an aspect of the meaning of the sentence, rather than of the “packaging” of the sentence meaning, and to relate to the speaker’s perception and construal of reality. Different topic constructions share the function of opposing a referring expression to the rest of the sentence in a more or less explicit way, excluding it from the scope of the illocutionary operator (assertion or others). They may be subsumed under three major templates grounded in the relation between: the event described in the sentence and an entity related to one of its participants (as in so-called hanging topic constructions), the event described in the sentence and its spatial or temporal location (as in constituent fronting), and the event described in the sentence and its primary participant (as in subject-changing constructions). Different degrees of conventionalization of the three templates account for topic and subject prominence across languages, European languages being notoriously subject prominent. In the discourse organization within individual languages, the different templates are associated with low degrees of activation of the topic referent or with topic-comment relations not inferable from the context.

Prominence given to some constituents and the information they carry in the unfolding of discourse is the key concept to all definitions of focus, critically surveyed by Jim Miller in the chapter on “Focus in the languages of Europe.” Prominence is therefore taken as the function allowing typological investigation in the broadest perspective, taking into account the various reasons for which constituents are made salient, be it introduction of new referents, contrast, or “exhaustive listing.” In this data-driven contribution, major attention is given to the grammatical means used for highlighting constituents on the basis of a sample of data drawn from various European languages, consisting of the oral instructions produced by two interlocutors in controlled conditions (the so-called “map task dialogues”), the responses to a questionnaire devised for this matter, and published corpora of spoken language. Written data are also considered. Major means found in the highlighting of constituents for different functions are clefts, particles, and rearrangement of word order. These means distribute differently on the diamesic dimension in some languages and furthermore, in a crosslinguistic perspective, they appear to concentrate in some areas of Europe, as in the case of clefts, characterizing Indo-European languages in the West of Europe (notably English), but unfound in Finno-Ugric and Turkic languages in the East of Europe.

Cohesion in discourse is explored by Yaron Matras and A. Machtelt Bolkestein (“Deixis and anaphora”) in a selection of different languages with respect to extratextual and intratextual reference, including anaphora in the proper sense. On the one hand, extratextual reference is distinguished by factors such as distance/proximity and visibility, discussed at length in the literature. On the other hand, factors relevant for the choice of means of expression for intertextual reference are the type of referent referred to (participant or text segment, such as the content of a clause) and the

degree of accessibility of the antecedent, mainly computed in terms of its distance from the anaphoric resumption, its syntactic status, and the potential competition with other referents. Distribution of deictic and anaphoric pronouns for intersentential anaphora in actual language use appears to result from the interplay of these factors: some favor, rather than oblige, the choice of a particular means of expression among the available ones and their effect may be better described in probabilistic terms on the basis of statistical considerations. Languages seem to differ considerably in the specific weight attributed to one or more factors in the selection of one form rather than another. Furthermore, pronouns used for extratextual reference appear also to be used for intratextual reference, but the relation between the two types of usage need further research in order to be accounted for in a straightforward way.

A particularly sensitive place in discourse is where new entities are introduced or new events are announced. Utterances in which these general functions are expressed are often said to be “all new.” Within a philosophical tradition going back to the work of Franz Brentano and Anton Marty in the late nineteenth century, this kind of utterance is said to convey a particular kind of judgment called *thetic*, that is, a judgment by which one affirms/negates what is being represented as a whole situation. In the last chapter of Part II (“*Theticity*”), Hans-Jürgen Sasse investigates five construction types for *thetic* utterances found in European languages and compares them from a variety of perspectives: text frequency, polysemy, restrictions imposed upon them by different languages, discourse functions, and lexical semantics. *Theticity* appears to be a crosslinguistically comparable phenomenon, although *theticity* itself is denied the status of a category, being rather a conglomeration of similar presuppositional/assertional conditions correlating with similar semantic areas. Furthermore, the *thetic*-categorical distinction appears not to be straightforward. Two of the comparable constructions found for *thetic* utterances – one with accented subject and one with verb-subject order – predominate and distribute over two larger areas covering the North-West and the South-East of Europe, respectively.

Part III (“From form to function”) comprises three contributions which investigate discourse functions of particular prosodic and (morpho)syntactic features.

Prosody is investigated by Alan Cruttenden (“The de-accenting of given information: a cognitive universal?”), with particular regard to discourse continuity and the intonational correlates of salience/non-salience of the pieces of information conveyed in a message. The supposed universality of the de-accenting of given information is investigated crosslinguistically by means of a repetition test in order to guarantee data comparability. In different setting-response dialogue types, the same lexical item is repeated in the setting and in the response, establishing a sequence of new-given information in the particular context of each test dialogue type. Test subjects for different languages were asked to read the translations of the original English setting-response pairs into their language and the data so obtained were evaluated by independent analysts. Prosodic treatment of the given items in differ-

ent languages in terms of optional and/or obligatory de-accenting and re-accenting appears to be influenced by the interaction of general discoursal and grammatical factors with language-specific variation. De-accenting and re-accenting are favored in most dialogue types in some languages (e.g., German and Spanish, respectively), while other languages do not show a clear tendency towards the one or the other kind of prosodic treatment (e.g., Italian and Swedish). As a consequence, de-accenting of given information cannot be claimed to be a simple cognitive universal.

An integrated view of the interplay of the syntactic and pragmatic levels of analysis with respect to basic constituent order and its typology is aimed at by Rosanna Sornicola ("Interaction of syntactic and pragmatic factors on basic word order in the languages of Europe"). Word order types found in the languages of Europe are considered under different perspectives: the interplay of position, syntactic function, and pragmatic function with respect to sentences with two arguments and one argument; the effect of the particular cases of verb and all-sentence focus; the influence of semantic features (animacy, definiteness, and referentiality of nominal arguments) and of the given-new dichotomy as a textual feature. The thorough survey of word order patterns of European languages according to these perspectives allows the establishment of a set of principles which set a network of conditions regulating the organization of word order patterns. The principles constrain the organization of the "sentence space," defined in terms of the relationship between constituents, constituent position, and the domain that contains them, barring, for example, the possibility for a constituent to occur in certain positions, even in marked constructions, as in the case of the space to the right of V in SOV languages. Particular attention is devoted to the interaction between constituency and the assignment of the focus function to O, that is, the constituent with the greater depth of embedding, discussed in detail with respect to the preverbal position of SOV languages, which could also result from the effect of semantic (e.g., animacy) or morphological (agglutination) features. This contribution opens a new perspective of typological relevance for the comparison of different word order patterns found in European (and non-European) languages.

Articles as a morphological device primarily devoted to the expression of definiteness and indefiniteness, that is, to signaling the referents' identifiability, are considered by Christoph Schroeder in the last chapter of Part III ("Articles and article systems in some areas of Europe"). The chapter surveys the articles and their systems found in languages of Northern Europe, of South-East Europe, and the Eastern margins of Europe in an areal perspective, taking into consideration the languages spoken in the interjacent areas with respect to other means used for the expression of (in)definiteness. The chapter contributes to the understanding of article systems that are significantly different from those known from the languages of Western Europe. Inventories with more than one series of definite articles serve to encode different types of reference. These might be dependent on whether the referent has already

been introduced in the actual discourse or is unique in a certain situational context, as, for example, in some Low German dialects, or else express specific distinctions with regard to the location of the referent in terms of restricted or wider situational contexts, as in Macedonian. The latter functional pattern seems to characterize some systems of articles deriving from possessive suffixes, as in Komi. In the areas of Europe considered in this contribution, the functional ranges of article uses reflect three continua of grammaticalization: from the numeral for ‘one’ to the indefinite article, from a demonstrative pronoun to a definite article, and from possessive suffixes of 2nd and 3rd persons to a definite article.

All of the contributions share a common functional-typological background and take into consideration the main factor contributing to the organization of linguistic material in discourse, that is, the opposition between spoken and written registers along the diamesic dimension, as discussed in Chapter 1. The diamesic dimension may play a crucial role in typological investigation of discourse organization. As, for example, Rosanna Sornicola points out in Chapter 7, standardized written registers of some European languages, for instance, German, may have been influenced by the literary tradition of classical languages and may show some discrepancies in word order with respect to spoken registers. Therefore, a careful treatment of these discrepancies is required in typology in order to arrive at reliable generalizations.

One major area of contention in the study of the pragmatic organization of discourse is represented by the definition of the relevant functions and of the terms adopted to refer to them. The discussions carried out by the EUROTYP group devoted to the “Pragmatic organization of discourse” in five years of common research have not been able to settle the questions pertaining to the definition of functions and the use of a common terminology. Therefore, it seemed wiser to avoid any kind of tentative standardization of terminology across the chapters of this volume. Some chapters discuss in a problematic way the major functions used in research on pragmatic organization of discourse, trying to arrive at a definition which reflects the author’s own vision of the matter: cf. Chapter 2 on topic, Chapter 3 on focus, and Chapter 4 on theticity. In the remnant chapters, the authors define the way in which they use the controversial terms (topic, theme, focus, all new), which may be in accordance, or only partly in accordance, with the terminological use found in other chapters. In any case, the readers won’t be puzzled by non-congruent usage, but rather will be helped in their orientation across a field of research in which consistency among different scholars is hard to arrive at. In this perspective, each chapter contributes different viewpoints on the same aspects of pragmatic organization of discourse, highlighting the difficult areas of descriptive and theoretical relevance and giving a complex picture of this field of study, which reflects its factual and phenomenological complexity.

This volume is the result of five years of common research during the EUROTYP program sponsored by the European Science Foundation, and of subsequent years of

elaboration and discussion of earlier versions of the chapters. Some chapters could be submitted earlier, some chapters could be submitted only later on, depending on the amount of exacting work requested by the treatment of their topics. Further unfortunate circumstances hindered the editorial work and the volume is now published with a considerable delay in the EUROTYP series.

The members of the group met on ten occasions between 1990 and 1994 in Pavia (Italy), Amsterdam (Netherlands), at “Il Ciocco” near Lucca (Italy), Bremen (Germany), Edinburgh (United Kingdom), San Sebastián (Spain), Bergamo (Italy), Naples (Italy), Le Bischenberg near Strasburg (France), and a last time again in Pavia (Italy).

The authors and I are grateful to all the institutions and the people who provided for a comfortable context of the meetings, and, in particular, the European Science Foundation for their organization and financial support. The authors and I are also grateful to the publisher Mouton de Gruyter and to the editors of the EUROTYP series, Bernard Comrie and Georg Bossong, for their support, help, and patience. The editing of the volume has received a fundamental contribution by the scrupulous revision of the co-editor Marcia L. Schwartz.

The members of the group on Pragmatic Organization of Discourse were, besides myself, Johannes Bechert (Bremen), A. Machtelt Bolkestein (Amsterdam), Alan Cruttenden (Manchester), M. M. Jocelyne Fernandez-Vest (LACITO, Paris), Elena S. Maslova (Saint Petersburg), Yaron Matras (Manchester), Jim Miller (Edinburgh), Jean Perrot (Paris), Hans-Jürgen Sasse (Cologne), Christoph Schroeder (Bremen), Rosannna Sornicola (Naples), and Barbara Wehr (Mainz).

Two members of the original group have left all of us: Johannes Bechert during the five years of common research, and A. Machtelt Bolkestein a few years afterwards. We all remember their scholarship and their contribution to the group work. This book is dedicated to their memory.

Spoken and written language

1. Introduction

The serious collection and analysis of spoken language corpora began in the 1960s, but little attempt has been made to relate the findings to theoretical linguistics.¹ In this chapter we argue that spontaneous spoken language differs in many respects from written language. The differences affect the general organization of discourse and all areas of syntax, from the elusiveness of sentences in spontaneous spoken language to the structure of noun phrases. The differences reach out from the core areas of syntax, morphosyntax, and discourse to areas such as historical change, language acquisition, and typology. We first show that the differences are extensive and deep and then use one major construction, relative clauses, to bring out the implications for typology. As will be argued in Section 8, the principal implication is that for a given language, the syntactic structures of spontaneous speech may fit one typological pattern, the syntactic structures of (formal) writing may fit another. Even more important, since this volume is concerned with discourse organization, is the fact that discourse organization is signaled by different devices in spontaneous speech and (formal) writing.

The comments on spoken English are based on two corpora. One is a corpus of 25,000 words of dialogue collected from pairs of informants as they were carrying out a task. The task involved drawing a route on a map, and the dialogues are referred to here as the “map task dialogues.” The second corpus contains 250,000 words of spontaneous conversation and is referred to as the “spontaneous conversations.” Data have also been taken from Macaulay (1991) and Milroy and Milroy (1993). The spoken Russian data is taken from Zemskaja (1973) and Lapteva (1976). The Sami data are from Fernandez-Vest (1987), the French and Finnish data from M. M. Jocelyne Fernandez-Vest’s database.

2. Distinguishing spoken and written language

2.1. Genres

The title of this chapter embodies a straightforward distinction between spoken and written language which is untenable. There are no grammatical or discourse differ-

ences that correspond with spoken vs. written texts. A first look at different texts quickly reveals that different genres must be recognized for both speech and writing: on the one hand, domestic conversation, dialogue in novels and plays, lectures, news broadcasts, discussion at academic conferences, legal speeches, sermons; on the other, personal letters, diaries, detective novels, poetry, academic monographs, “high” literature. And, of course, further distinctions can be drawn, say, within the category of high literature in English – the language of Jane Austen, the language of Charlotte Bronte, the language of Margaret Forster. Nonetheless, we will see below that spontaneous spoken language, especially conversation but also narratives and task-related dialogues, does have its own syntax and discourse organization.

2.1.1. Biber’s textual dimensions

Biber (1988) proposes a more subtle approach. He draws up a list of grammatical constructions and categories: for example, yes-no questions, IT clefts, WH-questions, agentless passives, 3rd person pronouns, adjectives. The list is established on the basis of what occurs with a high frequency in texts. On the basis of its frequency in types of text, each factor is assigned a weighting on a scale from +1 to –1. For instance, yes–no questions are frequent in conversations – what Biber (1988: 395) characterizes as texts produced under conditions of high personal involvement and real-time constraints. They are assigned a weighting of .79. Features that are not salient are assigned a low weighting: for the same type of texts, present tense has a weighting of .42 and a word length of .71.

On the basis of the factors and their cooccurrence, Biber (1988: 79–120) establishes various textual dimensions. The summary below gives the properties that cluster at either pole of each dimension.

1. Highly affective interaction and real-time constraints on language production vs. high informational content with time for editing
2. Narration of events vs. expository discourse
3. Explicit reference vs. situation-dependent reference
4. Overt signaling (for persuasive purposes) of the speaker’s/writer’s point of view, or overt assessment of the advisability and likelihood of an event vs. a lack of such signaling
5. Abstract, technical, and formal discourse vs. other types
6. Discourse that is informational but produced under real-time constraints vs. other types of discourse
7. The presence of hedges and qualifications vs. their absence

2.1.2. Spontaneous spoken language, written language, and education

The problem highlighted by Biber is that different text types line up differently with respect to each dimension, and a clear-cut distinction between written and spoken language fails to emerge. In the first paragraph of this chapter we referred to spontaneous spoken language, by which we mean language produced *impromptu* in relatively relaxed circumstances by a range of speakers – not all with higher education or even with much secondary education – talking about events in their own or other people's lives. The type of participants is crucial because of the effect of formal education and exposure to written language. Children in a literate society learn informal spoken conversational language as their native tongue. A proportion of children in, for example, Britain, listen to nursery rhymes, then short stories, then longer stories. From the age of five they are taught to read and write – not just to realize linguistic units as marks on paper, but to understand and use the structures and vocabulary of written English. This process lasts from age five to age sixteen at the very least and covers the language of personal narrative, description of scenes, reports of important public events, the language of modern and classical English literature, the description of experiments in science classes, the technical vocabulary and phraseology of mathematics, the sciences, modern studies, and so on.

For some speakers, the process of learning to use (as opposed to understand) all these different types of written English continues until they are eighteen, and through their years in higher education. Not everybody is equally capable of combining clauses into well-integrated sentences with subordinate adverbial clauses, participial phrases, and relative clauses introduced by a preposition plus a WH-word. Not everybody possesses the same range of vocabulary and the same skill at using their vocabulary accurately and effectively. (NB “not everybody” can apply to the set of university graduates, to the set of people who have had any kind of further education, to the set of people who have stayed at school till the age of eighteen, or to the entire population of the United Kingdom).

Biber's spontaneous spoken language data comes from the Survey of English Usage and was collected from middle-class, university-educated males. Many of the males were academics and among the segment of the population most affected by formal written language.² Investigators of spontaneous spoken speech in a number of countries (see references below) have discovered that such language typically has not just less complex structures but different structures; there are constructions that typically occur in spoken language but not in written language and vice versa. Biber maintains, correctly with respect to his data, that a clear-cut distinction between spoken and written language cannot be established. We claim that a clear-cut distinction can be established between typical spontaneous spoken language and typical formal written language: the constructions collected and described by the linguists listed below support that position. At the same time, we can subscribe to Biber's dimensions,

since the constructions that occur in spontaneous spoken language reflect properties such as situation-dependent reference, real-time constraints on language production, affective interaction, and so on.

At the time of writing this chapter, Longman had just circulated material advertising their *Grammar of Spoken and Written English* (Biber et al. 1998). This material incorporates an essential distinction between *repertoire* and *usage*. In one quote, Geoffrey Leech declares that spoken and written English have the same repertoire of constructions; in another quote, Douglas Biber declares that usage differs from situation to situation. The two quotes are quite compatible: in principle, Leech can be construed as saying that the same constructions are available to users of English whether speaking or writing. In practice, Biber can be construed as saying that users make different choices depending upon, *inter alia*, whether they are speaking or writing. Leech's assertion may be correct in principle but is actually unhelpful; *de facto* a good number of constructions that occur regularly in (spontaneous) speech do not occur in (formal) writing, and vice versa.

Let us emphasize again that what we focus on in this chapter is not just spoken language but

spontaneous spoken language – but note the discussion in Section 7 of just how far a planned speech can deviate from the written script if the speaker interacts with the audience rather than just reading out the script as though it were an unalterable text by some other writer. We believe that spontaneous spoken language deserves far more attention in linguistic research. It is what everyone acquires by the light of nature; it is what most people use most of the time; it is the source of much historical change. It is worth remarking here that the study of (spontaneous) spoken language is complicated by political and cultural attitudes. Consider the phenomenon of language variation. Variation that is mainly geographical is acknowledged, and linguists happily talk about, say, the ChiBemba dialect spoken in such and such a village or region. In contrast, variation between standard and nonstandard language brings problems which affect the analysis of spoken language because much of the study of spoken language has been based on nonstandard varieties or on varieties that differ from the standard written one. The difference in treatment is neatly reflected in the barbarous use of “substandard” instead of “nonstandard.” The approach taken here is that all spontaneous spoken language, whether standard or nonstandard, possesses certain syntactic properties: the sorts of structures observed in analyses of nonstandard English have parallels in the spoken standard Russian of professional people studied by Zemskaia (1973) and others.^{3,4}

2.2. Previous work on spoken and written language

2.2.1. General grammatical properties of spontaneous spoken language

The considerations outlined in Section 2.2 are important when we consider the opposing views that emerged from earlier work on spoken and written language. It is now generally accepted that spoken language, especially spontaneous speech, is very different from written language. The differences spring from various properties of spontaneous spoken language which are listed below – but the pragmatic nature of the properties does not mean that the syntax of spontaneous spoken language is to be treated as resulting from performance error (see Section 4).

- (i) Spontaneous speech is produced in real time, *impromptu* and with no opportunity for editing, whereas written language is generally produced with pauses for thought and with much editing
- (ii) Spontaneous speech is subject to the limitations of short-term memory in both speaker and hearer: it has been said (by the psycholinguist George Miller) that the short-term memory can hold 7 ± 2 bits of information
- (iii) Spontaneous speech is typically produced by people talking face-to-face in a particular context
- (iv) Spontaneous speech, by definition, involves pitch, amplitude, rhythm, and voice quality
- (v) Spontaneous face-to-face speech is accompanied by gestures, eye-gaze, facial expressions, and body postures, all of which signal information
- (i) and (ii) are reflected in five linguistic properties.
 - (a) Information is carefully staged, a small quantity of information being assigned to each phrase and clause
 - (b) Spontaneous spoken language typically has far less grammatical subordination than written language and much more coordination or simple parataxis
 - (c) The syntax of spontaneous spoken language is, in general, fragmented and unintegrated; phrases are less complex than phrases of written language; the clausal constructions are less complex
 - (d) The range of vocabulary in spontaneous language is less than in written language
 - (e) A number of constructions occur in spontaneous spoken language but not in written language, and vice versa

The simple nature of phrases, the unintegrated nature of the syntax, and the smaller range of vocabulary are all made possible by (iii), since typically a lot of information is shared or present in the situation of utterance and does not need to be articulated. Furthermore, a certain quantity of information can be signaled by the ancillary systems mentioned in (iv) and (v).⁵

2.2.2. General discourse properties of spontaneous spoken language

Fernandez (1982: 259–260) draws attention to the use of repetition and inversion of word order in spontaneous speech as strategies of persuasion (and, we might add, attention-holding). She gives a good example of a stylistic scheme combining /repetition + iconic cohesion + circular cohesion/ from Finnish spoken by Samis:

- (1) Finnish
 Alavieskassa ja Taivalkoskella | olen ollut.
 Alavieska-LOC(intern.) and Taivalkoski-LOC(extern.) I-have been
 Olen ollut / Nuorgamissa . . . Palokoskella
 I-have been Nuorgam-LOC(intern.) Palokoski-LOC(extern.)
 Karigasniemessä | olen ollu Rautuskaijissa |
 Karigasniemi-LOC(intern.) I-have been Rautuskaidi-LOC(intern.)
 olen ollu ja joka paikassa | olen ollu
 I-have been and every place-LOC I-have been

The first clause contains coordinated NPs in the locative case (internal/inessive, or external/adessive), in initial position before the verb. In the second clause the locative case NPs are in final position, simply juxtaposed (listing). The remaining three clauses all have the order locative case NP – verb; the NP in the final clause sums up and generalizes the locative case NPs – *joka paikassa* ‘every place.’ The boundary between the locative case NPs and the verb is further marked by two non-verbal signals. The first is gestural – a slap of the speaker’s hand on the table. The second is intonational: in each utterance, apart from the second one, which has the structure theme – rheme (*I have been / in Nuorgam . . .*, rising pitch / falling pitch), the portion to the left of the | carries falling pitch, whereas the portion to the right carries level pitch. The portions with the falling pitch are the rhematic part of the clause, and the portions with the level pitch are analyzed as post-rheme elements or mnemes (Fr. *mnémème*, see Fernandez-Vest 1994: 197 ff.)

Consider the following two English texts.

- (2) then he said why was I always trying to CHANGE him and I said proba-
 bly because he’s such an obnoxious thoughtless selfish overbearing self-
 righteous hypocritical arrogant loudmouthed misogynist bastard
- (3) Mr. X was tall bald hair to here and a beard he looked like Jesus he liked
 to think he did he liked to think he was Jesus he was horrible the most
 horrible man

(2) is taken from a humorous postcard (the words are spoken by one woman to another as they sit at the kitchen table chatting over a glass of wine and a cigarette) and the humor comes precisely from the fact that in such a relaxed setting, most people cannot produce even two or three well-chosen adjectives off-the-cuff, far less a

string of nine, and even the most fluent speakers are hard pressed to produce complex syntax at the right moment in stressful situations. Most of us suffer from *l'esprit de l'escalier*. (3) is taken from spontaneous conversation. It illustrates the combination of verbal and nonverbal communication, since *hair to here* is accompanied by a gesture – Mr. X is bald on top but has hair almost down to his shoulders. The remainder of (3) demonstrates how in spontaneous speech speakers make an assertion and then repeat it, not necessarily exactly but with additional information. Thus, the next clause carries the assertion that Mr. X looked like Jesus, while the following two clauses each carries the qualification that Mr. X liked to think he looked like Jesus. The penultimate clause conveys the judgment that Mr. X is horrible, while the final noun phrase intensifies the judgment; Mr. X is not just horrible but most horrible. (The change of syntax is to be expected; [*Mr. X is*] *most horrible* is unlikely to occur outside formal written English. *Mr. X is the most horrible man* sounds more natural, although the complement noun phrase has the air of a ready-made unit.)

2.2.3. Analyses contra a special status for spontaneous spoken language

Some scholars have taken the opposite view that spontaneous spoken language has complex syntax. Poole and Field (1976) found spoken discourse to have a significantly greater degree of subordination, elaboration of syntactic structure, and use of adverbs. Halliday (1989: 76–91) maintained that written language is complex in that highly compact and simple syntactic constructions are loaded with many lexical items. Spoken language is complex in a different way, having intricate syntactic structure with a considerable proportion of subordinate clauses. Lexical items are spread over these subordinate clauses, reducing the lexical density of each one. Halliday illustrates the distinctions via the written sentence *The use of this method of control unquestionably leads to safer and faster train running in the most adverse weather conditions* and a possible spoken variant ||| *If this method of control is used || trains will unquestionably (be able to) run more safely and faster || (even) when the weather conditions are most adverse* ||| (Halliday 1989: 79). Unfortunately, various features mark this utterance as written or enunciated by someone who speaks like a book: the impersonal conditional clause, the adverb *unquestionably*, the noun phrases *method of control* and *weather conditions*, and the adjective phrase *most adverse*. A more plausible utterance would be: *If you control the trains this way they'll definitely run safer and faster – in really bad weather and all*. Even Halliday's (1989: 79) "more natural" spoken version contains sequences such as *you can be quite sure that . . . , no matter how bad the weather gets, than they would otherwise*, which are quite untypical of spontaneous spoken language.

As Beaman (1984: 51) suggests – and cf. Biber's dimensions – it is important to distinguish between the modality (spoken vs. written) and register (formality). She proposes that what have been treated as differences between spoken and written dis-

course may in fact reflect differences in formality or planning time, and to these should be added the property mentioned in Section 2.2, exposure to higher education. Poole and Field gathered their data from undergraduate students in a formal setting, while the data cited by Halliday appears to have been invented by him and unfortunately reflects the complex nature of the spoken language of academics.⁶

3. Sentences and clauses⁷

3.1. Introduction

The position defended here is that the sentence should be regarded as a low-level discourse unit of written language, that clauses and phrases are units of both spoken and written language, and that, as suggested by Halliday (1989), sequences of clauses in spoken language may form clause complexes, which do not have the structural integrity of sentences. Indeed Sornicola (1981) argues convincingly on the basis of Neapolitan Italian that much spontaneous language does not even have a syntactic structure in which phrases combine into clauses or clauses into integrated clause complexes. Rather, the structure consists of blocks of syntax (phrases) with little or no syntactic linkage and requiring from the listener a larger than usual exercise of inference based on contextual and world knowledge. Sornicola demonstrates that such fragments should be treated as the structures that speakers aim to produce in spontaneous speech, and not as the remnants of clauses that have fallen apart as a result of performance errors.⁸ Sornicola's approach seems to deprive formal models of any solid unit for the analysis of interclause relations, reference, and anaphora, but alternative frameworks are now available, albeit relatively undeveloped, in the theories of rhetorical structure and discourse representation.

3.2. System sentences and text sentences

We will assume a major distinction between language system and language behavior. The language system consists of the syntactic, morphosyntactic, semantic, phonological, and graphological principles controlling the generation of semantic and syntactic structures, the insertion of lexical items into the syntactic structures, and the realization of the structures as speech or writing. The products of speaking and writing are texts, which may be spontaneous or deliberately elicited by investigators. Hypotheses about particular language systems or the general nature of language systems are based on texts and intuitions. It is essential to distinguish units that can be recognized in texts from units that belong to the hypothesized language system. For instance, when a modern written text in English (or French, German, etc.) is examined, the analyst finds the text divided into units whose initial boundary is signaled

by a capital letter and whose final boundary is signaled by a full stop. These units, text sentences, are determined by the author of the text – if the text is a personal letter or an essay, the division into text sentences typically remains unchanged once the text has left the author's hands, but if the text is, say, a book being prepared for publication, a subeditor may question some of the author's decisions. For example, a particular author may have a subordinate clause of concession or a relative clause constituting a whole text sentence, whereas many copy editors prefer to include at least one main clause in each text sentence.

In order to handle the patterns of constituent structure and dependency relations that manifest themselves during the analysis of texts, linguists establish units called sentences. These are what were referred to above as system sentences. System sentences do regularly map on to text sentences in written texts but not always. For instance, restrictive relative clauses are treated in linguistic analyses of English as embedded in noun phrases containing the nouns they modify. This enables the analyst to handle the dependency relation between relative clause and the noun it modifies, to capture what is indicated by distributional evidence – (determiner), noun and relative clause form a single constituent – and to compose the denotation and reference of that single constituent. In modern written English texts, however, it is not unusual to find a text sentence consisting entirely of a relative clause, while the noun it modifies is in another text sentence: *The door was opened by a man. Who appeared to be about seven foot tall and six foot wide.* Analysts have two approaches to choose from. They can allow their grammar to generate a single system sentence in which the relative clause is embedded inside the oblique object of the main clause, or they can allow their grammar to generate two separate system sentences: one the main clause and the other, the relative clause. The former approach allows dependency relations and compositionality to be handled straightforwardly, but requires the structure to be dis-integrated for mapping onto the surface structure. The latter does not require dis-integration but does call for devices to handle the dependencies and compositional relations that cross the sentence boundary.

Strictly speaking, the different units should be clearly kept apart by means of different terms, such as “text sentence,” “system sentence,” “text clause,” “system clause,” etc., as in Lyons (1977). Here, “sentence” and “clause” will be used where it is clear from the context whether the unit belongs to text or to the language system.

What text units can be recognized in spoken language? In written language, sentences and clauses (and phrases, paragraphs, etc.) are obvious in any text laid-out according to the conventions of the society in which it was written. Interestingly and importantly, the relevant conventions differ from society to society and from one period of time to another in the same society; the organization of clauses into sentences differs from Jane Austen to A. S. Byatt, and from British newspapers to French newspapers – but text sentences are clearly delimited in all these genres. The question of what units can be recognized in spoken language and are useful for its analysis is

not so easily resolved. Some analysts maintain that sentences are not recognizable in spoken language, and others maintain that they are.

System sentences are postulated by linguists in order to handle distribution and dependency relations, and must be retained if this goal is to be achieved. The status of system sentences in the language system of written English is not disputed, and is usually regarded as self-evident because of the clear delimitation of text sentences. It is not usually remembered that sentences are learned through the process of reading and writing, and are *taught* to the majority of language users, whereas clauses are acquired without specific teaching. Children in the early stages of primary school typically produce single-clause sentences and have to acquire the ability (partly by instruction, partly by reading) to combine a number of clauses into a sentence. Written texts are produced, and analyzed, not by the untutored, but by people who have been inducted over a long time into the conventions governing the organization of written texts in their society.

The central problem is that it is far from self-evident that the language system of spoken English has sentences, for the simple reason that text sentences are hard to locate in spoken texts. Clauses are easily recognized: even where pauses and a pitch contour with appropriate scope are missing, a given verb and its complements can be picked out. Of course, one reply to the objection is that the system sentences employed by linguists need not correspond to text sentences. Against this, it can be argued that system sentences do not map onto text sentences in spontaneous language, because system sentences are based on the prototype concept of a sentence as containing at least one main clause and possibly other coordinated main clauses and/or subordinated clauses. (That this is indeed the prototype concept is easily verified by examining popular manuals such as Burton [1986], the literature in any generative framework or the training offered to school pupils). It can also be argued that in the language system, the essential locus of both dependency relations and distributional properties is in fact the clause (see Section 3.4).

If spoken texts lack sentences, the language system must be analyzed as having clauses combining into clause complexes, as suggested by Halliday (1989). There are two major types of syntactic relationship: embedding and combining. Adverbial clauses only combine, that is, they are not part of any constituent in a matrix clause. Only relative and complement clauses can be embedded, since relative clauses are regularly part of an NP and complement clauses function as arguments to verbs. However, many occurrences of relative clauses cannot be treated as embedded, especially if, in dialogue, they occur in a different turn from the head noun and come from a different speaker. Relative clauses do occur as the sole constituents of sentences even in written English, although such syntactic arrangements might not be considered good style. English is not alone, since the same phenomenon is found from Classical Latin to Modern English. A French example, albeit of a special kind of relative construction, is given in (4b)(iv) and (v) below.^{9,10}

3.3. Sentences in spoken texts

3.3.1. Introduction

This section briefly surveys the arguments for and against the sentence as a unit in spoken texts and as an analytical unit in accounts of spoken language. The case against has been stated most clearly by Halliday (1989: 66), who argues that the basic unit of syntax is the clause. Clauses occur singly or in complexes, and clause and clause complex are indispensable concepts for the study of both spoken and written syntax. Sentences in written language developed from the desire to mark clause complexes; the initial capital letter of the first word in a clause complex, and the full stop following the final word signal which clauses the writer wants the reader to construe as interconnected. Of course, clauses are also interconnected in spoken language; the difference is that interconnectedness is not signaled by adjacency nor even by the relevant clauses occurring in the same turn (in conversation) or under the same intonation contour (in narrative).

A number of researchers recognize the problematic nature of the sentence in spoken language. Quirk et al. (1985: 47) state that the sentence boundaries can be difficult to locate “particularly in spoken language”¹¹ and point out that the question “What counts as a grammatical English sentence?” does not always permit a decisive answer. They deal with the difficulty by avoiding any definition of sentence while continuing to use the term for a unit greater than the clause. Linell (1988: 54) reaffirms the lack of clear-cut sentences in spoken language and adds that talk consists of phrases and clauses loosely related to each other and combining into structures less clear and hierarchical than the structures dealt with in grammar books. Similar points had been made earlier by Brown et al. (1984: 16–18). Brown et al. and Linell are apparently satisfied to work with phrases and clauses, precisely the position adopted here.¹²

Sentences in spoken language are defended by Chafe and Danielewicz (1987: 94–96). They invoke “prototypical intonation units,” consisting of a single coherent intonation contour, possibly followed by a pause and stretching over a maximum of six words, which often constitute a clause but which may also constitute a phrase or simply a fragment of syntax. Chafe and Danielewicz (1987: 103) further say that speakers appear to produce sentence-final intonation when they judge that they have come to the end of some coherent content sequence. One difficulty with Chafe and Danielewicz’s account is that their sentences correspond more to short paragraphs than to the prototypical written sentence. The intonation contours they describe may encompass one or more main clauses, not conjoined but simply adjacent to each other. Conversely, the same type of intonation contour may encompass a mere phrase.¹³

3.3.2. Intuitions about sentences in spoken language

Wackernagel-Jolles (1971: 148–169) demonstrated that speakers do not share intuitions about what counts as a sentence in spoken language. She got groups of thirty to fifty final year undergraduate students at a German university to listen to recordings of narratives by native speakers of German. (The narratives had been prompted by questions from an interviewer). Each text was played through once to allow the students to accustom themselves to the speaker's voice. They were then given a transcription of the recording, without punctuation. The text skipped to a new line only where there was a change of speaker. The recording was then played through a second time, and the students were asked to draw a line in the text wherever they thought a sentence ended. Agreement as to sentence endings ranged from 13 out of 20 in one text, to 6 out of 29 in another. The former text was the telling of a fairy tale; the latter, a panel-beater recounting his early life and his war experiences. Wackernagel-Jolles (1971: 149) comments that uninterrupted story-telling was especially conducive to clear intonation signals but that no correlation emerged between speed and clarity of pronunciation and degree of agreement. Speakers/writers, who as university students can doubtless organize their own written texts into acceptable sentences, were unanimous about the final boundary for less than half of the sentences in the texts.¹⁴ For them the sentence is a relatively fluid unit.

3.3.3. The sentence: A changing concept

The view of the sentence as a relatively fluid discourse unit in written language fits with the fact that, in written English, text sentence boundaries vary from one historical period to another, from one genre to another, and from one individual to another. Moreover, text sentences vary from one language to another. In contrast, clauses (and phrases), which are central units of syntax, are not subject to such cultural variation and rhetorical fashion.¹⁵

As observed in Section 3.2, the concept of text sentence is not stable across cultural boundaries and can be manipulated to achieve particular stylistic effects. For instance, the French weekly *L'Express* has a house style that encourages phrases and subordinate clauses of all types to be presented as single text sentences, as exemplified in (4).

- (4) a. (i) Ils sont de bonne foi.
they are of good faith (i.e. sincere)
- (ii) Comme étaient de bonne foi ces ménagères engueulant
as were of good faith these housewives shouting-at
les refuzniks
the refuseniks

- b. (i) Certains ont invité les contestataires à aller
certain ones invited the objectors to go
- (ii) se plaindre auprès de Raïssa Gorbatchev.
to-complain to Raïsa Gorbachev
- (iii) Avec quelques commentaires grossiers sur la femme
with some comments coarse on the wife
- (iv) du secrétaire général.
of-the secretary general
- (v) Ce qui n' était pas très difficile.
which not was not very difficult

(4a)(ii) contains a subordinate adverbial clause of comparison constituting a complete sentence, namely *Comme étaient de bonne foi ces ménagères engueulant les refuzniks*. (4b)(iii) has a prepositional phrase as a separate sentence, while (4b)(v) has a relative clause as a separate sentence. The unusual segmentation, a strategic chunking of the information which imitates the rhythm of conversation, creates suspense and, as the writer probably hoped, the appropriate reaction from readers.

The Russian weekly *Argumenty i Fakty* offers the examples in (5).

- (5) a. Komandir soznatel'no idet na risk.
(the)-captain consciously goes on risk
'The captain consciously takes a risk.'
- b. Nadejas', što peregruzki ne budet.
hoping, that overloading not will-be

In (5b) *nadejas'* is a nonfinite verb form, a gerund. Equally interesting examples are in (6).

- (6) a. Tol'ko za poslednij god ob"em aviaperevozok v obščem
Only in (the) last year (the) volume of air-journeys in general
po SNG sokratilsja na 30 %.
in (the) CIS has fallen by 30 %
- b. Na Ukraine i v Rossii – na tret', v Tadžikistane –
In (the) Ukraine and in Russia – by (a) third in Tadžikistane –
na polovinu.
by (a) half

The interesting point about (6b) is that it is a complete sentence consisting of a gapping construction, and the constituent required in order to interpret the gap, *sokratilsja*, is in the previous sentence. (6b) is relevant not only as an example of a verbless sequence functioning as a sentence in a text, but as an example of a dependency carrying over from one sentence to another. This property means that in a generative analysis, the gap must be handled by a mechanism that can operate across

sentence boundaries as well as across clause boundaries, and the example is another piece of evidence in support of the view that the abandonment of sentences as analytical units with respect to spoken language does not create new problems but requires mechanisms already required by existing problems.

Further problems are posed by dialogue. Consider the examples in (7)–(8) from the map task dialogues.

- (7) B: right if you go from the front giraffe's foot about
hold on let me see
if you go down about straight down about 6 cms
you find the waterhole
and it's a big hole . . . with reeds round the side of it . . . and
animals drinking out of it
and it's about
it's a an oval hole
it's about 2 cms wide north to south
and from the side to side it's about – 3 cms wide . . .
- (8) A: you go down to the bridge
B: uhuh to the left of the swamp?
A: to the left of the swamp – taking a gentle curve southwest

(7) illustrates how in spontaneous spoken language information is carefully staged, in the sense of being spread out over different clauses. Most of the clauses are simple clauses and are simple in structure, though one clause has two prepositional phrases with a participial phrase inside one of the latter – *with reeds round the side of it* and *(with) animals drinking out of it*. It would be possible to gather the clauses into sentences, but various possibilities are open. For example, we might decide to have *You find the waterhole and it's a big hole . . . [and it's about]. It's an oval hole. It's about 2 cms wide north to south. And from . . .* Another possible version is *You find the waterhole. And it's a big hole . . . It's about 2 cms wide north to south and from . . .*

The basic difficulty is that in collecting the clauses into sentences, we rely on our ability to recognize clauses and on our knowledge of the stylistic conventions for written dialogue. As in (1), the intonation and pause boundaries do not coincide with the possible sentence boundaries, and to add to the difficulties, the prepositional phrases *with reeds round the side of it* and *(with) animals drinking out of it* are separated from the initial part of the clause and from each other by a long pause. It is in fact unclear whether these chunks should be analyzed as combining into a single clause. The analyst can combine the clauses into sentences, but the combining process is arbitrary and the sentences would not contribute to the analysis of the data as a coherent text. Coherence relations (say, as part of a discourse representation theory) must apply to clauses and indeed phrases, and sentences are not necessary.

Similarly, (7) could have been *written* as the compact, dense, syntactically integrated piece of prose in (9), but the characteristics of (9) cannot be invoked as criteria in the analysis of (7). It would be rather like taking a piece of written language, rewriting it, analyzing the rewritten piece, and presenting the analysis as pertaining to the original passage!

- (9) It's a big oval waterhole about 2 cms wide north to south and about 3 cms wide from side to side, surrounded by reeds and with animals drinking out of it.

(8) exemplifies another relationship that cannot reasonably be analyzed by invoking a single sentence. The free participial phrase *taking a gentle curve southwest*, modifies the clause produced in a previous turn by the same speaker. Participial phrases are discourse-dependent in the sense that the listener cannot interpret them without reference to a previous piece of text; at the very least, a subject has to be found for the participle itself. The nearest candidate for the subject is in the first line of (8), but this is not a reason for analyzing the participial phrase as belonging to one and the same sentence as *you go down to the bridge*. Note that speaker A was not interrupted in the process of producing a single sentence. *You go down to the bridge* is a completed utterance, with appropriate intonation. B signals acceptance of the instruction with *uhuh*, looks at the map and realizes that he needs more information: *to the left of the swamp?* Speaker A produces the participial phrase in response to speaker B's question.

The development of sentence structure in written language is discussed, for example, by Palmer and Guiraud. Palmer (1954: 119) remarks that complex sentences in written Latin prose were consciously developed by generations of writers and that the resulting body of rhetorical conventions had to be taught. Once the vernacular Romance languages had broken away from Latin, the organization of clauses into sentences had to be established for each vernacular Romance language as it began to be used as a vehicle for prose literature. Guiraud (1963: 113) observes that the Old French literary language was very close to the spoken language, having an essentially paratactic organization of clauses into larger units. Such a syntax “n’a jamais eu l’entraînement ou la pratique qui l’auraient pliée à l’expression d’une pensée élaborée; elle ignore l’articulation logique de la démonstration scientifique ou les méandres de l’argumentation philosophique” [has never had the training or practice that would have formed it to the expression of elaborated thought; it is unaware of the logical structure of scientific argument or the meanderings of philosophical discussion, translation by Jim Miller]. It is significant that Guiraud mentions the uses to which language is put by literate human beings; French syntax did not develop a complex, hypotactic organization of clauses by some mysterious process but through the conscious efforts of certain literate people to convert French into an instrument suitable for the purposes served by Classical Latin.¹⁶

It must be emphasized that neither Palmer nor Guiraud suggests that the development of subordinate clauses was subsequent to the development of written Latin and written French. Rather, they allude to the organization of several or many clauses into a sentence and the way in which the conventions governing this organization were developed by the users of a given written language. They also allude to the special development of language by scientists for the accurate description of scientific data, experiments, and theories, which is quite in accord with the comments by Ong (1982) on the role of the Royal Society in Britain in the late seventeenth century in encouraging the emergence of a special scientific language and the logical presentation of hypotheses, data, and conclusions.

3.4. Sentences and syntactic analysis

3.4.1. The clause as the central unit of syntactic analysis

We turn now to sentence and clause in linguistic analysis. The burden of the preceding discussion is that the sentence is not a unit that can be recognized in spoken texts or applied in their analysis. In contrast, the sentence is a prominent unit in written texts and requires a corresponding analytical unit. However, there is evidence that the clause should be taken as the major locus of distributional and dependency relations and not the (system) sentence.

Crystal (1987: 94) provides a concise rendering of the definition of “sentence” provided by Bloomfield (1935: 170): “a sentence is the largest unit to which syntactic rules apply.” Interestingly, Bloomfield’s sentences each consist of a single finite clause. Bloomfield does treat the problem of two or more juxtaposed main clauses without pauses or intonation break by invoking a set of pitch phonemes, but this analysis is decisively countered by Matthews (1981: 30–34), on the grounds that intonation is continuous, the phonemic principle of sameness vs. distinctness does not apply, and there are no rules governing parataxis.¹⁷

In any case, the syntactic units (in spoken or written language) affected by the rules of distribution and dependency relations are the phrase and the clause. The clause is the locus of the densest dependency and distributional properties, although a few dependency relations cross clause boundaries, and, in written language, a few dependency relations cross sentence boundaries. That dependency relations cross clause boundaries could be interpreted as supporting the sentence as an analytical unit even in spoken language, but this is counteracted by the fact that dependency relations cross text sentence boundaries in written language. Because dependency relations cross text sentence boundaries, a complete grammatical theory must have a mechanism for specifying such dependencies, and whatever the mechanism is, it will undoubtedly be able to specify dependencies from clause to clause when the clauses are gathered, not into a sentence, but into a text.

Cross-clause dependencies are frequent in Classical Greek, as exemplified in (10)–(14). Let us note first that, like Russian, the densest network of dependencies is inside the Classical Greek clause. This is demonstrated in (10a), where the verb *akoúousi* assigns nominative case to its subject noun and accusative case to its direct object noun. (In contrast, in (10b) the verb *khṛṛtai* assigns dative case to its object noun, and the very assignment of dative case raises the question whether that verb takes a direct object or an oblique object.) In turn the nouns spread case, number, and gender to any dependent articles and adjectives: *ándres* assigns nominative, plural, and masculine to *hoi* and *kakoí*; and *lógous* spreads accusative, plural, and masculine to *toús*. *Gunaikós* (of the woman) is a feminine noun in the genitive case because of its relationship to *lógous*. *Gunaikós* spreads genitive, singular, and feminine to *tês* and *sofês*.

- (10) a. *hoi* *kakoí* *ándres* *ouk*
 the-NOM.PL.M evil-NOM.PL.M men-NOM.PL.M not
 akoúousi *toús* *tês* *sofês*
 listen-to-PRES.3PL the-ACC.PL.M the-GEN.SG.F wise-GEN.SG.F
 gunaikòs *lógous*
 woman-GEN.SG.F word-ACC.PL.M
 ‘The evil men are not listening to the words of the wise woman.’
 b. *hē* *gunē* *khṛṛtai* *toís*
 the-NOM.SG.F woman-NOM.SG.F use-PRES.3SG the-DAT.PL
 biblíois
 books-DAT.PL
 ‘The woman is using the books.’

As in English, complement-taking verbs in Classical Greek control the complementizer in the complement clause. Verbs of saying take *hóti* ‘that,’ as in (11), verbs of movement take *hína* ‘in order to,’ as in (12), verbs of inquiring take *ei*, as in (13), and certain specific verbs take *hópōs*, as in (14).

- (11) a. *légei* *hóti* *gráfei*
 he/she-says that he/she-is-writing
 b. *eîpen* *hóti* *gráfoi*
 he/she-said that he/she-was-writing

 (12) a. *érkhetai* *hína* *ídē*
 is coming in-order-that sees
 ‘He/she is coming to see.’
 b. *élthen* *hína* *ídoi*
 came in-order-that sees
 ‘He/she came to see.’

- (13) a. *punthánetai ei akóousi toûs lógous*
 asks if hear the words
 'He/she is asking if they hear the words.'
- b. *epútheto ei akóoien toûs lógous*
 asked if hear the words
 'He/she asked if they heard the words.'
- (14) a. *spoudázei hópōs akóúsetai toûs lógous*
 hurries to hear the words
 'He/she is hurrying to hear the words.'
- b. *espoúdase hópōs akóúsetai toûs lógous*
 hurried to hear the words
 'He/she hurried to hear the words.'

Note that although *hópōs* in (14a) and (14b) has been translated with 'to,' the complement clause is finite, *akóúsetai* being third person singular future. In addition to verbs selecting complementizers, there is another dependency crossing the clause boundary. When the verb in the matrix clause is past tense, the verb in the complement clause is in the optative mood. When the verb in the matrix clause is present tense, the verb in the complement clause is either indicative, after *hóti* in (11a) and *ei* in (13a), or subjunctive, after *hína* in (12a). When the verb in the matrix clause is aorist, the verb in the complement clause is in the optative mood.¹⁸ The exception to these changes in mood is *hópōs* in (14a) and (14b), which requires the verb to be future tense. In the last case it is *hópōs* and not the verb in the matrix clause that governs the occurrence of future tense.

Dependencies do not cross clause boundaries into adverbial clauses. Certain combinations of adverbial clause and main clause appear to involve cross-clause dependencies, such as the rules in Classical Greek governing clause combinations expressing fulfilled or unfulfilled conditions as in (15).

- (15) a. *ei toûto epoíoun, ēdíkoun*
 if this they-were-doing they-were-wrong
- b. *ei toûto épraxen, ēdíkēsan án*
 if this they-had-done they-would-have-been-wrong particle

The English copula + adjective structure corresponds to a single verb in Greek. (15a) expresses a fulfilled condition; the conditional clause contains an imperfect verb, *epoíoun*, as does the main clause, *ēdíkoun*. (15b) expresses a remote, unfulfilled condition. The conditional clause contains an aorist form, *épraxen*, and the main clause contains an aorist form with the particle *án*. Such examples, however (both the Classical Greek ones and their English equivalents), are not instances of dependencies crossing from clause to another. The syntactic constraints affect both the main and adverbial clauses, and the dependencies appear to be associated with the entire combination, rather than flowing from the main clause to the adverbial clause.

The above examples of cross-clause dependencies do not vitiate the proposition that the clause is the site of the densest network of dependencies. In each of the above examples, at most two dependencies cross the clause boundary, the selection of complementizer by the verb and the selection of mood in the complement clause. Inside each clause is a much greater number of dependencies. The verb controls the type of constituents it requires; some verbs allow two NPs, others allow three, and yet others allow only one; some verbs allow adjective phrases or PPs. In a given clause the verb assigns case to the dependent nouns and controls the choice of preposition in the PPs. Inside the phrasal constituents the head, N, A, or P, assigns case to its dependent constituents, and a head N also assigns gender and number. And there may be further PPs inside the NP and AP. Not all these dependencies flow directly from the verb, but they are all sited inside a given clause.

3.4.2. Dependencies crossing sentence boundaries

To close this discussion of dependencies, let us consider the Russian text in (16), exemplifying dependencies crossing text sentence boundaries.

- (16) a. Ètot portnoj byl krasivo starejuščij mužčina
this tailor was handsomely growing-old man
b. šest' večerov v nedelju stojal on za stolom,
six evenings in week stood he at table,
c. rezal šil proglašival švy utjugom.
cut sewed smoothed seams with-iron.
d. Zarabatyval den'gi. Voskresen'e provodil na ippodrome
earned money. Sunday spent at racecourse

In (16) the first sentence has the full masculine singular NP *ètot portnoj*; the second sentence has the masculine singular pronoun *on*; and the third and fourth sentences have the subjectless masculine singular verbs *zarabatyval* and *provodil*. The properties “masculine” and “singular” are projected by *ètot portnoj* into the pronoun in the second sentence and then into the verb forms in the third and fourth sentences.

3.4.3. Sentences, clause, and distribution

With respect to distribution, it is equally obvious that the classic distributional criteria for constituent structure apply within clauses rather than sentences. For example, in a recent introduction to transformational grammar (Radford 1988: 69–75), the vast bulk of the distributional evidence relates to single main clauses. Where there is more than one clause, one reason is that the additional material is needed to provide a convincing linguistic context; for example, in *Down the hill John ran, as fast as he could* the adverbial clause of manner, *as fast as he could*, lends naturalness to the

fronted prepositional phrase *down the hill*. A second reason is that the extra clause is a relative clause or a complement clause, that is, clauses that are embedded inside arguments of the verb in the main clause. One example is *He explained to her all the terrible problems that he had encountered*, where the relative clause gives the necessary weight to the final noun phrase.

The only clauses that have distribution inside a unit bigger than the clause are adverbial clauses. In written English and in relatively formal spoken English, adverbial clauses of time and reason, for example, can precede or follow the main clause with which they combine, but in informal spoken English, they tend to follow the main clause. That is, since their distribution even in written English is limited, they are no more than mild exceptions to the rule that the clause is the main focus of distributional properties. In any case, it is not clear that even adverbial clauses can be moved inside a sentence in written English. The difficulty is that not all subordinate clauses are equally subordinate, where subordination is measured in terms of possible constructions and word orders. Main clauses permit a large range of constructions – declarative, interrogative, imperative, tag questions – and a large range of word orders, whereas subordinate clauses vary in the extent to which they allow constructions other than declarative and word orders other than subject – verb – direct object. Consider the examples in (17).

- (17) a. Because Aunt Norris came into the room, Fanny stopped speaking.
 b. Fanny stopped speaking because Aunt Norris came into the room.

In (17a) the adverbial clause of reason *because Aunt Norris came into the room* is at the front of the sentence and has limited structural possibilities. *Into the room*, for example, cannot be moved to the front of the clause, and the subject NP and the verb cannot be transposed, but these changes can be carried out in the *because* clause in (17b). Compare (18a) and (18b).

- (18) a. *Because into the room came Aunt Norris, Fanny stopped speaking.
 b. Fanny stopped speaking, because into the room came Aunt Norris.

It has been suggested that *because* clauses have a different discourse function in sentence-initial and sentence-final positions (Chafe 1984). In sentence-initial position, *because* clauses (indeed, adverbial clauses in general) function as a guide to information flow, whereas in sentence-final position, adverbial clauses simply add something to the assertion conveyed by the main clause. In sequences of main clause – *because* clause, the two clauses almost function like coordinate clauses. The *because* clause may relate to one particular constituent in the main clause or may not relate directly to the main clause at all. That is, the un-subordinate nature of such *because* clauses, as evidenced by their syntactic flexibility, is accompanied by semantic flexibility.¹⁹ For instance, in *Fiona isn't coming to work today because her husband phoned up to say she was ill*, the *because* clause presents, not the reason

for Fiona's absence, but the reason for the speaker being able to state that Fiona is not coming in to work. For some speakers, adverbial clauses of reason following the main clause can even contain interrogative structures: for example, *I'm not going to the party because who's going to be there?*

Complement clauses are *prima facie* better examples of distribution inside the sentence, given that there appears to be no major difference in meaning between a sentence such as (19a), with a complement clause in sentence-initial position, and (19b), with the complement clause in sentence-final position.

- (19) a. That the enemy was approaching the town apparently did not worry the inhabitants.
 b. It apparently did not worry the inhabitants that the enemy was approaching the town.

For written English it is indeed correct that complement clauses can occur at the beginning or end of sentences, but in spontaneous spoken English examples like (19a) are practically unknown. There are none in the corpora of conversation and map task dialogues, and they are very rare even in formal spoken language such as is heard in serious discussion programs on radio and television. That is, in spontaneous spoken English, complement clauses are not mobile but fixed.

To sum up, there is very little evidence to support either text sentences or system sentences in spontaneous spoken language. Planned or semi-planned spoken language is different, but is typically heavily influenced by the units and organization of written language. Much of the language system of a given language is medium-independent, but some is dependent, most obviously the complex syntactic constructions and the vocabulary that are typical of written language but not spoken language. Equally, there are constructions and vocabulary that occur in spontaneous spoken language but not in written language. The system differences can be kept to a minimum by appealing to the notion of discourse rules specific to a given medium. The discourse rules for written language map one or more clauses into sentences where appropriate; the discourse rules for spontaneous spoken language do not.

4. "Syntactic fragments," competence, and performance

Sornicola (1981) and Enkvist (1982) argue strongly that even apparently fragmented syntax should be treated on its own terms and not as the degraded realization of an ideal clause or clause complex, particularly as it is not always obvious what the ideal structure might be. Consider (20), from spontaneous spoken English.

- (20) A: whose idea // was it
 B1: Charlie Richardson's
 C: uhuh

- B2: they got one of the teachers that we always play jokes on / one of the young women / they got her to write it

The sequence *one of the young women* is spoken on a much lower pitch than the surrounding material, and there is a clear tonic on *we* and *write* in B2. What syntactic structure in her competence was the speaker aiming at? It might have been *they got one of the teachers that we always play jokes on to write it*, but the inserted explanatory material *one of the young women* has disturbed the flow of the surface syntax. An alternative is that *they got one of the teachers that we always play jokes on* corresponds to one originally complete syntactic structure and that *they got her to write it* corresponds to another complete piece of syntax. That is, although the inserted material can be seen as interrupting the syntactic performance, it is equally possible that the text clauses correspond to the original abstract syntactic structures that the speaker had in mind.

The same difficulty is posed by spontaneous spoken Russian. Consider (21), from Zemskaja (1973: 27).

- (21) a. moloko raznosit / ne prixodila eščë?
milk she-delivers / not she-came yet
'The woman who delivers the milk, has she not come yet?'
b. u okna ležala / kapriznaja očën'
at window she-lay / moody very
'The woman in the bed by the window was very moody.'

The "missing" syntax in (21) is not just a noun but a noun and a relative pronoun. By "missing" is meant that in writing (21a) and (21b) have to be converted to (22) and (23).

- (22) ženščina, kotoraja moloko raznosit, ne prixodila eščë?
woman who milk she-delivers not she-came yet?
(23) ženščina, kotoraja u okna ležala, kapriznaja očën'
woman who at window she-lay moody very

The essential point is that (21a) and (21b) must on no account be thought of as reduced versions of (22) and (23). Rather, the written examples have a very different clausal structure from the spoken ones. (22) and (23) consist of a main clause in which the subject NP contains a relative clause: *ženščina, kotoraja moloko raznosit* and *ženščina, kotoraja u okna ležala*. (21a) and (21b) consist of two main clauses juxtaposed: *moloko raznosit* and *ne prixodila eščë* in (21a) and *u okna ležala* and *kapriznaja očën'* in (21b). From the perspective of English, one possibility that suggests itself is a headless relative construction. Could the Russian examples be equivalent to *Has who delivers the milk come yet?* and *Who was in the bed by the window*

was very moody? The attraction of the English examples is that, in spite of their infelicitous syntax, they provide subject NPs for the main clauses, namely *who*. This attraction does not transfer to the Russian examples, principally because the Russian equivalents of English headless relatives involve either a general noun such as *čelovek* followed by a relative clause, or a correlative construction: for example, *I will live where you live*, with the headless relative *where you live*, corresponds to the Russian example in (24).

- (24) Ja budu žit' tam, gde ty budeš' žit'
I will live there where you will live

The correlation is between *tam* and *gde*. The Russian equivalent of *the one who* is *tot, kto*. In order to analyze (21a) and (21b) as headless relatives, we would have to postulate a structure with a relative clause containing both an empty relative WH-NP operator and modifying an empty head NP. Since this structure provides no anchoring referent for the WH-NP operator, an analysis in terms of headless relatives lacks appeal.

The analysis proposed above – that (21a) and (21b) consist of juxtaposed main clauses – is made possible by the frequent lack of subject NPs in spoken Russian. One of the striking features of Zemskaja's data is the frequency of zero subject NPs and zero direct object NPs, especially in conversation. Even in written Russian an entity introduced by an overt NP in one clause can be referred to by zero subject NPs over five or six clauses, and even across sentence boundaries. It is clear that the speakers of (21a) and (21b) were referring to entities already mentioned or were treating the entities as highly given, and the absence of overt subject NPs is normal in this context. For (21a) an appropriate gloss is *she delivers milk – has she come yet*, and for (22b), *she was in bed by the window – she was/is very moody*. The occurrence of the feminine verb forms *prixodila* and *ležala* indicates that the speakers were referring to specific persons, women in both cases.

A further example of unintegrated syntax that is typical of spontaneous spoken English is given in (25) from the map task dialogues.

- (25) what you're going to do – you're going to go up past the allotments.

Weinert and Miller (1996) were taken to task by one referee for confusing competence and performance in their discussion of the structure of WH-clefts in English. According to the referee, (25) is the result of performance factors. But what would be the structure shared by (25) and the integrated WH-cleft of written English in (26)?

- (26) What you are going to do is go up past the allotments.

The written construction has a copula but the spoken construction has not. The spoken construction has a second main clause with progressive aspect but the written

construction has not. The spoken construction has an overt subject NP in the second clause but the written construction has not. A single integrated source structure would require a large number of unusual empty categories. Of course, perhaps the referee simply meant that speakers start to produce an integrated WH-clause, run into planning problems, cut the construction short after the WH-clause, and produce a main clause *faute de mieux*. This interpretation does not square with the fact that the context surrounding the WH-clefts in the map task dialogues and the spontaneous conversations display no symptoms of planning problems such as hesitations or repetitions. It also ignores the fact that the structure in (25) is easily analyzed as an information packaging instruction in the sense of Vallduví (1993). The WH-clause encodes an instruction to the listener to erase the information in the file “next action” and to prepare to enter fresh (and correct) information.²⁰ (For the arguments supporting this analysis, see Weinert and Miller 1996) The WH-word points forward to some entity, whether proposition or concrete individual, which the speaker specifies in the second clause.

Fernandez-Vest (1987: 686 ff.) observes that discourse particles play an important role in the presentation of disconnected phrases as a coherent message. In the Northern Sami example in (27) – an orally transmitted language until the 1980s – the particles connect chunks of syntax that do not by themselves make up a clause conveying a proposition, but correspond in fact to a dialogical style inserted in the monological narrative. (The chunks encoded with an exponent letter are in bold.)

(27) Sami

[The anecdote is about *láttánat* (‘the landmen,’ i.e. non-Sami people), who enjoy fishing and wandering in the mountains, but freeze to death with the first drops of rain, as they are unable to start a fire.]

ORAL VERSION

Muhto maid^a dat^b dákkar^c / báikegoddálaš boah^tá gi lea ollu mehc^{iid} johtán^d / dathan^{ef} gal^g arvinge^h fidne dola galⁱ. Na^j i das^k / mihkkige^l go dat^m lea dola ožžonⁿ dat^o dat^p gal^q i jáddat / dan gal^r i ajibeavvisge^{s+} (...)

‘But **what^a then^b such a^c** / local guy arrives **who has a lot in the forest wandered^d / he^e certainly^f yes^g even^h** in rainy weather⁺ / gets a fire **yesⁱ. Well^j in this^k / nothing no^l** when **he^m has fire-gotⁿ he^o of course^p sure^q** does not put it out / **for sure^r never^{s+}** (...).’

WRITTEN VERSION

Muhto go boah^{tá}’ ollu mehc^{iid} johtán^d báikegoddálaš, sonhan^{e’f} gal^g fidne arvinge^h dola. Go son^{m’} lea dola ožžon dola^{n’}, de⁺ láttán^{o’} i jáddat dan ollege^{s’} (...)

‘But **when arrives^{a’} a much-in the forest-having-wandered^{d’}** regional guy, **he^{e’} certainly^{f’} yes^{g’}** gets **even^{h’}** in rainy weather a fire. When **he^{m’} has got a fire^{n’}, then⁺ the landman^{o’}** does not put it out **at all^{s’}** (...).’

The interlocutive dimension, omnipresent in the Sami oral discourse, is exemplified here by the interrogative *maid* ‘what,’ responded in the following sentence by an elliptic comment – *Na* (opening particle ‘well, now’) *i das mihkige* ‘in this nothing **no**.’ The informant chose automatically to delete these dialogical signs (as well as many other discourse particles) in the written version.

The final words in this section come from Heath (1985). He was writing about another phenomenon, but his sentiments are appropriate to the study of spoken language in general:

There has been a recurrent tendency in much syntactic research to distinguish between an underlying, rather crystalline “grammar,” which then interacts in real speech with a distinct outer “psycholinguistic” component, the latter being especially concerned with short-term memory limitations, linear ordering of major clause constituents, resolution of surface ambiguities, etc. My view is that these two aspects of language are far more tightly welded to each other than it seems at first sight. (Heath 1985: 108)

5. Clause and clause complex

A major difference between spontaneous spoken language and written language lies in the organization of clauses into clause complexes. The typical relationship between clauses in the former is hypotaxis, while in the latter, clauses are related both hypotactically and paratactically.

Schulz (1973: 19–50) draws attention to the paratactic expression of causal relationships between clauses, pointing out that such relationships by no means call for subordinate clauses but can be expressed by particles, as in (28) and (29).

- (28) Meistens, nachmittags, geh ich dann mit den Kinder raus,
 mostly in-the-afternoon go I then with the children out
 die müssen ja auch frische Luft haben.
 they must particle particle fresh air have
 ‘I usually go out with the children in the afternoon. . . because after all
 they must have fresh air.’

- (29) Da haben die nichts mit verdient.
 there have they nothing with earned
 ‘They earned nothing there.’
 Die Kumpels, die hierher kamen, die hatten doch wenig
 the lads who here came they had particle little
 Geld.
 money
 ‘My mates who came here had little money.’

Schulz glosses both examples by means of clauses introduced with *weil* ‘because’ and with the finite verb in clause-final position, the classical mark of subordinate clauses in German: *Meistens gehe ich nachmittags mit den Kindern nach draußen . . . , weil sie frische Luft haben müssen* and *Damit verdienten sie nichts, weil die Kumpels, die hierher kamen, wenig Geld hatten*. In the actual spoken examples the particles *auch* and *doch* signal that there is a relationship between the two clauses.

Spontaneous spoken Russian exhibits a similar lack of hypotaxis, and even a lack of particles as functioning in (28) and (29). Lapteva (1976: 305–321) adduces many examples in order to demonstrate that clauses which are clearly in a discourse relation of subordination are regularly not in a relation of syntactic subordination.²¹ Examples are given in (30).

- (30) a. Ja ne uspel. Ja prišel tam uže ne
 I not was-in-time I arrived-PFV there already not
 prinimali.
 accept-IPFV-3PL
 ‘I was not in time. When I arrived there, they were no longer accepting
 [people, applications, . . .].’
- b. Vy uezžali iz Moskvy doždik byl?
 you were-traveling out-of Moscow rain-DIM was?
 ‘When you were leaving Moscow, was it raining a little?’
- c. čto mne delat’ slesarja ne dozovešsja?
 what I-DAT do-INF joiner not call-2SG.REFL
 ‘What am I to do if you can’t get through to the joiner?’
- d. K. včera xvastalsja novuju palatku kupil.
 K. yesterday boast-IPFV.PST new tent buy-PFV.PST.3SG
 ‘K. was boasting yesterday that he had bought a new tent.’

The key property of the above examples is that the relationship between the clauses is not signaled by means of complementizers or particles. Even the notion of discourse subordination is not obviously applicable. In an English example such as *When you left Moscow, was it raining?* the adverbial time clause *when you left Moscow* can be interpreted but the interpretation is obviously incomplete; the event of leaving of Moscow is being used as a point of orientation for another event, but this other event is not specified. In (30b), however, *vy uezžali iz Moskvy* could be taken as a main clause, as could *doždik byl*.²²

Fernandez-Vest (1994: 95–96), however, points out that intonation can signal the integration of two clauses as opposed to simple coordination or juxtaposition. Consider the French examples in (31).

- (31) a. T'auras pas de dessert [↓] t'es pas venu avec
 you won't have any dessert (because) you didn't come with
 nous →
 us
- b. Il n'a pas plu [↓] le linge est sec →
 it didn't rain (since) the washing is dry

A special pitch pattern signals that *t'es pas venu avec nous* and *le linge est sec* are not free-floating clauses but are linked to the first clause in each example. Of course, the type of semantic link – time, reason, concession – does have to be reconstructed by the addressee but the information structural function is clear: a rhematic segment (falling intonation) is followed by a post-thematic one (flat intonation).

To conclude this section, we note briefly that certain constructions occur in written language but typically do not occur in spontaneous spoken language. The examples in (32) are constructions from English.

- (32) a. Subordinate clauses introduced by *although, since, as*
 b. WH relative clauses, especially structures with the WH-form inside a PP or with a clause or clauses inside the relative clause.
 c. Gapping:
 Sue likes, and Bill hates, crosswords.
 Celia likes Van Gogh, and Bill – Rembrandt.
 d. Full gerunds:
 His having the book at all astonished me.
 e. Accusative + infinitive:
 I considered her to be the best candidate.
 f. Participial phrases:
 Sitting at the window, I noticed a car at the bank.
 Covered in confusion, he apologized.
 g. Certain conditional constructions:
 Were you to write to her, she would forgive you.
 Should you meet him, pass on my best wishes.
 h. Indirect questions:
 This is not just a case of whether the two words can combine.

Some of the constructions above will be picked up in Section 8 on typology.

6. Noun phrases

Spontaneous spoken language and written language differ with respect to the structure of noun phrases. For English, the difference lies in the complexity of the phrase; noun phrases with more than one modifier, in addition to a determiner, are rare, and

the modifier is typically a single adjective or a simple prepositional phrase.²³ In other languages, however, different structures occur in spoken language. Zemskaja (1973: 254–255, 391) cites examples such as those in (33).²⁴

- (33) a. Tam est' kraby banočka odna
 there be-3SG crab-NOM.PL tin-NOM.SG one-NOM.SG
 'There's one tin of crabs there.'
- b. Modnyj kupili emu kostjum
 fashionable-ACC.M.SG buy-PFV.PST.3PL he-DAT suit-ACC.M.SG
 'They bought him a fashionable suit.'

In written Russian we would expect to find *odna banočka krab*, with the genitive form *krab*. Zemskaja comments that *kraby* in (33a) is informationally important, hence the different constituent order. In (33b) *modnyj* modifies *kostjum* but is separated from it by the rest of the clause. Although there is no space here to discuss the matter in detail, it should be noted that a major question about constituent structure arises – namely, should the sequences *kraby banočka odna* and *modnyj kostjum* be treated as deriving from solidary noun phrases or should they be analyzed as consisting of separate chunks – for example, *kraby* and *banočka odna* – that are simply juxtaposed? Whatever the answer, it will certainly involve theories of lexical entries and the selection of anaphors, and appeals to the concept of scrambling are not likely to be helpful.²⁵

7. Discourse

It is perhaps surprising that spontaneous spoken language and written language should differ so much with respect to sentences, and to type and complexity of clause construction. The differences in discourse organization should be less unexpected; conversation, with turn-taking and continuous interaction between speaker and addressees, is quite different from written genres, except of course imitations of conversation in plays and novels. Planned spoken performances such as narratives, lectures, and political speeches have a general structure that runs parallel to the structure of written narratives, lectures, and speeches. The structure may well not be signaled in the same way. For instance, the narrator of an oral tale may mark a major change of event, place, and characters with a long pause and by a phrase such as “While all that was happening, . . .” or the classic cliché “Meanwhile, back at the ranch . . .” Written narratives do not necessarily contain such continuity phrases, but major changes are signaled by gaps of different lengths and by the beginning of a new section or a new chapter.

Speakers and writers do share certain tasks; they have to open texts by introducing the entities they want to talk or write about and the situations in which the entities participate; they have to signal the continuity of entities in the discourse by using

shorter NPs to refer to them, including null NPs; they have to introduce new entities to the original stock; when appropriate, they have to make entities more prominent or salient (NB “prominent” and “salient” are not used here as technical terms) than the other entities referred to in the text; and they have to indicate when a portion of text continues a general topic, changes the general topic, or expands or contradicts what has just been said or written. The interaction between speaker and addressees involves the speaker in tasks such as making sure that the addressees pay attention to the right entity at the right time, and checking whether the addressees follow the message and whether or not they agree with it.

Formal written texts, such as novels, and especially expository texts, such as academic monographs and legal judgments, typically reflect adherence to general conventions such as beginning at the beginning of the matter and signaling continuity or change of topic, agreement, or contradiction, by the use of adverbial words and phrases such as *furthermore*, *in contrast*, *nevertheless*, *while it is the case that* ..., and so on.

The difference between spoken and written texts is clearly brought out by the comparison of written versions of papers at an academic conference with the spoken versions recorded as they were delivered (see Fernandez-Vest 1994: 144–158). The spoken versions were not given off the cuff but were anchored to a written version. One section of the published written version is given below.

- (34) la difficulté fondamentale (insurmontable?)
 the difficulty fundamental (insurmountable?)
 en traduction automatique
 in translation automatic
 ‘The fundamental (insurmountable?) difficulty in machine translation’
 tient à ce qui fait la spécificité de tout texte –
 holds to that which makes the specificity of all text
 littéraire ou non –
 literary or not
 ‘lies in what constitutes the peculiar nature of any text – whether literary
 or not –’
 écrit dans une langue,
 written in a language
 à savoir la formulation elliptique (...).
 to know the formulation elliptic (...)
 ‘written in a language, namely ellipsis (...).’
 Or un automate ne peut travailler que sur l’explicite (...).
 now a machine not can work but on the explicit (...)
 ‘Now a machine can only handle what is explicit (...).’

Le paradoxe est de devoir réconcilier le texte et l'automate.
 The paradox is to have-to reconcile the text and the machine
 'The paradox is having to bring together text and machine.'

The spoken text is 30 % longer than the written text. The speaker consulted his notes in order to keep to the logical order of topics and in order to provide correct documentary references, but the content and shape of the spoken text was determined by his interaction with the audience. As Fernandez-Vest (1994: 150) observes, the speaker's presentation, unless it sticks strictly to its written version, does not require special analysis of the language of conference talks; the interaction with the audience has its roots in the practice of everyday conversation. The spoken text is given in (35).

- (35) a. alors/ la difficulté fondamentale / en fait / en traduction
 well the difficulty fundamental in fact in translation
 automatique
 automatic
 'Well, the basic difficulty in machine translation really'
- b. elle est due à ce qui fait la spécificité de tout texte
 it is due to that which makes the specificity of all text
 écrit dans une langue naturelle /
 written in a language natural
 'is due to what constitutes the specific nature of any text written in a natural language'
- c. que ce texte soit littéraire ou NON /
 whether this text is literary or not
 'whether a given text is literary or not'
- d. et qu' on a évoqué / à plusieurs reprises / pendant
 and which one has invoked at several times during
 ces deux journées /
 these two days
 'and which has been mentioned several times during the past two days'
- e. c'est à dire ce qui fait la spécificité d'un texte en
 that-is to say that which makes the specificity of-a text in
 langue naturelle
 language natural
 'that is to say what constitutes the peculiar nature of a natural language text'
- f. c'est sa formulation elliptique
 that-is its formulation elliptic
 'is ellipsis'

- g. Hein/ le problème de L'ELLIPSE /
PART the problem of the-ellipsis
'OK the problem of ellipsis'
- h. la formulation elliptique avec le texte en langue naturelle /
the formulation elliptic with the text in language natural
'elliptical formulation with natural language text'
- i. la LINÉARITÉ / laconique / de la surface (...)
the linearity laconic of the surface (...)
'the laconic linearity of the surface text'
- j. le texte n'étant autre chose que /
the text not-being other thing than
le SIGNE le POINTEUR hein /
the signe the pointer PART
'the text being simply the sign the pointer OK'
- k. qui suggère le sens plutôt qu'il ne le détaille.
which suggests the meaning rather than it not it detail
'which suggests the meaning rather than specifying it in detail.'

The NP specifying the new topic in line (a), *la difficulté fondamentale en traduction automatique* is not part of a clause; the following clause is separated from the NP by a short pause and has its own subject, *elle*, which picks up the text-initial NP. The NP is preceded by the particle *alors* signaling that the speaker is about to start; the first and important subpart of the NP, *la difficulté fondamentale*, is followed by another particle, *en fait*. This particle conveys the speaker's knowledge that various major problems afflict machine translation, but that he is about to mention a difficulty which he regards as the most fundamental. At the same time, the particle makes *la difficulté fondamentale* stand out. In line (c) the phrase *littéraire ou non* in the written text is expanded into a full clause repeating the noun *texte* to indicate continuity of referent and giving accentual prominence to NON. Line (d) ties in the about-to-be-mentioned fundamental difficulty to discussions at the conference: the difficulty is nothing bizarre but is well-known and the audience can relax. (Fernandez-Vest observes that the audience consisted of people who were linguists but not specialists of machine translation.) Line (e) is a second run-up to the mention of the fundamental difficulty, which is specified in line (f). The difficulty is respecified, with alterations, in lines (g)–(i). These lines are introduced by the particle *hein*, spoken with a rising pitch and followed by a short pause. The second occurrence of *hein*, in line (j), gives prominence to the important words SIGNE and POINTEUR, and the significance of these words is spelled out in line (k).

Other major differences in the discourse organization of spoken and written texts relate to the devices by which constituents, and the information they carry, are highlighted. We must make clear here and now that "highlighting" is not intended as a technical term. We are interested in any devices by which one constituent is made

to stand out from the surrounding text: using the expressions with their everyday meaning, we say that constituents are highlighted, made salient or made prominent. We will avoid “focus,” which is a technical term with multiple uses which will be discussed in Chapter 3. Suffice it to say here that constituents can be highlighted because the speaker wants to mention an entity for the first time, reintroduce an entity already mentioned, or contrast one entity with another. Highlighting can be achieved by special word order, special constructions, particles, and of course pitch and volume. We will not be concerned with the latter two, although they are key properties of spoken language and are excluded from written language except insofar as a writer chooses to use italics or bold font or capital letters to hint at suprasegmental modulations. The highlighting devices discussed below consist of syntactic constructions or particles that are typical of spoken language – the examples are English, French, and Russian – but untypical of and excluded from written language.

A common construction in spoken English (indeed, in the spoken varieties of many languages) has an NP followed by a complete clause containing a pronoun that picks up the referent of the NP. Examples from spontaneous conversation in English are given in (36).

- (36) a. it's not bad – ma Dad he doesn't say a lot
 b. the driver he's really friendly – you get a good laugh with him
 c. well another maths teacher that I dinnae get he must've corrected my papers

(36a) is a reply to the question *what do you get discipline-wise* which is part of a conversation about bringing up children. (36b) is a reply to the enquiry *what are the people like to work with – the drivers and that?* (36c) is part of a discussion about the speaker's examination results. Note that the independent NPs in (36a) and (36b) are very simple, just a determiner and a noun, and that there are no indications of planning problems such as hesitation or repetition. (36c) does contain a complex NP, but this example also offers no indications of planning problems. The resumptive pronouns in the examples above are mostly subjects, with the exception of the oblique object *with him*. The pronoun can have any grammatical function; an example of a direct object is *it* in *the book I lent you have you finished it yet?* It is more likely that the construction is connected with the spreading of information over syntactic constituents in small doses, with the need to highlight entities being introduced into the discourse, and with the general constraint that constituents are kept simple in spontaneous spoken language. Of course, in (36b) *the driver* is given; it links back to a preceding utterance *what are the people like to work with – the drivers and that* but this specific driver is being mentioned for the first time.

It is worth pointing out that the independent NP is not just a result of fronting; the link between the NP and the resumptive pronoun does not obey constraints such as the “complex NP constraint.” Consider (37).

- (37) The new boss did you hear the rumor that he's leaving already?

Russian examples from Zemskaja (1973) are given in (38) and (39).

- (38) a. a rebjata èti / kotorye igrajut v komandax / skol'ko
but fellows these who play in teams how-many
im let?
to-them of-years
'But these guys that play in the teams – how old are they?'
b. A kak ja uznaju kotoraja že Tat'jana?
but how I will-recognize which-one particle Tat'jana
'But how will I recognize which one is Tat'jana?'
c. Tanja / u nee belaja šapočka s pomponom.
Tanja at her white hat with pompom
'Tanja has a white hat with a pompom.'

Russian also has a construction in which the independent NP is not picked up by a resumptive pronoun. Instead the independent NP sets a frame, and the NPs in the following clause relate to items in the frame. Consider (39).

- (39) a. ved' ržanoj xleb / vot ètot zapax specifičeskij
you-know rye bread FOC that smell specific
ostalsja s detstva.
has-remained since childhood
'You know, rye bread, that particular smell has stayed with me since I was a child.'
b. Sobaka / vseгда poly grjaznye.
dog always floors dirty
'When you have a dog/with a dog the floors are always dirty.'
c. Deti / bez šumu ne obojdeš'sja.
children without noise not you-will-manage
'Where there are children, you can't avoid noise.'

Fernandez-Vest (1995) offers the French examples of a categoric (negative) general assertion (40) tempered by a more personal and specific one – detached topic + pragmatic particle (41) and the interesting example in (42) with two independent NPs.

- (40) C'est pas le cas de tous les petits vieux.
It is not the case of all the little old-men
'It is not the case with all the little old men.'
(41) Mon grand-père / bon / ça ne l'amusait pas tellement.
my grandfather PART that not him-amused not so
'My grandfather, well, it didn't amuse him all that much.'

- (42) Ta frangine, sa bague, c'est du toc.
 your sister her ring it's of-the fake
 'Your sister's ring, it's fake.'

(42) could be rendered in (spoken) Scottish English as *See your sister, see her ring, it's fake*. It could be rendered in other varieties of spoken English, including the standard, as *You know your sister's ring . . .* or *You see your sister's ring . . .* These are special constructions for the first mention of entities that are known to the speaker and hearer.

Very common highlighting devices in spoken English are *I mean/do you mean* and *like*. The latter is exemplified in (43).²⁶

- (43) A: er I'm I'm not very sure ++ what I'm supposed to be doing
 B: em and then you have to go down again
 A: **like** I go past the collapsed shelter?

As mentioned above, both written and spoken English have WH-clefts, but the typical spoken WH-cleft differs from the written WH-cleft. The written one has a WH-clause followed by the copula, followed in turn by an infinitive phrase or a bare verb stem, as in *What she did was to cut all the cuffs off his shirts/what she did was cut all the cuffs off his shirts*. The typical spoken cleft has the WH-clause followed immediately by a main clause, as in (44). The two clauses can have the same or different subjects.

- (44) no you had no food attached you got your meal hours of course but //
 what you did in the evenings you carried a / sandwich or two / and you
 had a little break in between
- (45) what you're doing you're going up past the market garden
- (46) I'll give it a little stir because what happens – things tend to settle a bit

Spoken English also has a range of alternatives to the WH-cleft which do not typically occur in written texts.

- (47) A: What about Edinburgh do the people go up there
 B: oh yeah a lot
 C: oh aye especially at night they go to the pictures but **the thing is** if
 you go to the pictures if you go to the late show you're you've to
 run for buses
- (48) A: **that's the bad thing about the halls of residence** there's always
 people knocking on your door
- (49) A: thing is he's watching the man he's not watching the ball
 B: right I see so **is that the idea of this then?** so you go straight to
 where I am instead of going round the picket fence?