

Mastering English



Topics in English Linguistics

22

Editor

Herman Wekker

Mouton de Gruyter
Berlin · New York

Mastering English

An Advanced Grammar for
Non-native and Native Speakers

Carl Bache
Niels Davidsen-Nielsen

Mouton de Gruyter
Berlin · New York 1997

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

Bache, Carl, 1953–
Mastering English : an advanced grammar for non-native
and native speakers / Carl Bache ; Niels Davidsen-Nielsen.
p. cm. – (Topics in English linguistics ; 22)
Includes bibliographical references and index.
ISBN 3-11-015535-4 (alk. paper)
ISBN 3-11-015536-2 (pbk)
I. English language – Grammar. I. Davidsen-Nielsen,
Niels. II. Title. III. Series.
PE1106.B27 1997
428.2'4–dc21 97-23493
CIP

Die Deutsche Bibliothek – Cataloging-in-Publication Data

Bache, Carl:
Mastering English : an advanced grammar for non-native and
native speakers / Carl Bache ; Niels Davidsen-Nielsen. – Berlin ;
New York : Mouton de Gruyter, 1997
(Topics in English linguistics ; 22)
ISBN 3-11-015535-4
NE: GT

© Copyright 1997 by Walter de Gruyter & Co., 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
permission in writing from the publisher.
Printing: WB-Druck GmbH, Rieden am Forggensee.
Binding: Lüderitz & Bauer, Berlin.
Printed in Germany.

Acknowledgements

In 1993, the Danish Research Council for the Humanities launched five foreign-language grammar projects. In being commissioned to write the grammar of English in this programme, we were largely relieved from our teaching duties. In addition to being sponsored by the research council, this book has received financial support from Odense University (on the recommendation of Fritz Larsen and Carl-Erik Lindberg), the Copenhagen Business School, Ingeniør N. M. Knudsens Fond and from Ib and Elise Bache. For all this generous support we would like to offer our sincere thanks.

Peter Harder (Copenhagen University) and Fritz Larsen (Odense University) have been our readers during the entire project and have provided us with extremely helpful and perceptive criticism. We are also grateful to Alex Klinge (Copenhagen Business School), Leo Høye (Odense University) and Henning Kirkegaard (Odense University) for proposing a number of highly constructive revisions. Without the liberal assistance of these five scholars the task of completing this book would have been harder and the result poorer. Needless to say none of these persons can in any way be held responsible for remaining errors and obscurities in the text.

We would also like to thank our colleagues in the parallel grammar projects on French, German, Italian and Russian for commenting critically on selected topics in our grammar, presented at a series of very inspiring cross-linguistic symposia.

In our project we have been fortunate enough to have Christian Heyde Petersen (Odense University) as our research assistant. One of his various tasks was to compile the subject index. We would also like to thank Stefan B. Andersen (Odense University Press), Kjeld V. Sørensen, Jørn Erik Wennerstrøm and Elsebeth Jensen (Odense University Computer Service Centre) for technical assistance. For assistance received in the late phases of the work our thanks go to Anne Marie Køllgaard (Copenhagen Business School) for proofreading the entire manuscript and completing the subject index, and to Eva Bang and Mimi Swiatecka (students at the Copenhagen Business School) for helping out with the word index.

Our grammar has been tried out on students of English at Odense University and at the Copenhagen Business School, not only by ourselves but also by Alex Klinge, Christian Heyde Petersen and Marianne Stølen. Our thanks are due to these colleagues and to our students (particularly Stefan Mogensen) for providing a great deal of valuable feedback.

When in doubt about the acceptability of our examples we have conferred with native speakers of English, in particular John Murphy (Copenhagen Business School), John Dienhart, Leo Hoyer, Sharon Millar, Tom Pettitt (Odense University) and Penny Rosier and Phil Staines (Sydney). Our thanks go to these colleagues as well for responding orally – with patience and apparent interest – to questions posed over a long period, often at inconvenient hours.

Copenhagen and Odense, April 1997

Carl Bache & Niels Davidsen-Nielsen

Contents

Part I

Chapter 1: Preliminaries	1
1.1. Goals and framework	1
1.2. Organization	1
1.3. Data	2
1.4. Varieties of English	3
1.5. Standard English	4
1.6. Grammatical variation	5
1.7. Variation according to medium	7
1.8. English for Special Purposes	11
 Chapter 2: An introduction to syntax	 13
2.1. The word	13
2.2. The sentence	15
2.3. Grammatical structure	18
2.4. Linearity and the principle of proximity	19
2.5. Constituency	20
2.6. Linguistic creativity and ambiguity	22
2.7. Competence and performance	24
2.8. Syntagmatic and paradigmatic relations	25
2.9. Recapitulation	28
 Chapter 3: Elementary sentence analysis	 29
3.1. The basic form and function approach	29
3.1.1. Form and function	29
3.1.2. Sentence functions	29
3.1.3. Four form types	31
3.1.4. Word classes	32
3.1.5. Simple complete analyses	36
3.1.6. Discontinuity	37
3.2. Sentence functions and sentence structures	38
3.2.1. The predicator	38
3.2.2. The subject	41
3.2.3. Basic sentence structures	44
3.2.4. More sentence structures	46
3.2.5. The direct object	47

3.2.6. The indirect object	50
3.2.7. The subject complement	52
3.2.8. The object complement	54
3.2.9. Adverbials	55
3.3. The internal structure of complex forms	59
3.3.1. The group	59
3.3.2. The preposition group	63
3.3.3. The compound unit	66
3.3.4. The clause	68
3.3.5. Embedding	69
Chapter 4: Advanced sentence analysis	71
4.1. Stacking	71
4.1.1. Form stacks	71
4.1.2. Function stacks: predicates and predication	73
4.1.3. A final comment on stacking	77
4.2. Missing constituents, ellipsis and pro-forms	78
4.2.1. The zero convention	78
4.2.2. Types of ellipsis	81
4.2.3. Pro-forms	83
4.3. Complex predicators	85
4.3.1. Phrasal verbs	85
4.3.2. Prepositional verbs	88
4.3.3. Phrasal-prepositional verbs	93
4.3.4. Other complex predicators	95
4.4. The top of the tree	97
4.4.1. Exit Sent	97
4.4.2. Communicative functions	98
4.4.3. The forms of communicative functions	102
4.4.4. Non-declarative clauses	104
4.4.5. Block language	105
4.5. Vocatives, interjections and dislocation	106
4.6. A final word on advanced sentence analysis	110

Part II

Chapter 5: Constituent order	111
5.1. Introduction	111
5.2. Functions of constituent order	112
5.2.1. Grammatical relations	112
5.2.2. Illocutionary value	112

5.2.3. Information structure	113
5.3. Inversion	116
5.3.1. Preliminaries	116
5.3.2. Full and partial subject-predicator inversion	116
5.3.3. Partial inversion with illocutionary value	118
5.3.4. Partial inversion caused by an initial constituent	120
5.3.5. Full inversion and information structure	123
5.3.6. Full inversion after a fronted adverbial	124
5.3.7. Full inversion after a fronted participial predication stack	125
5.3.8. Full inversion after a fronted subject complement	127
5.3.9. Full inversion after a fronted object	127
5.3.10. Inversion of other sentence constituents	130
5.4. Constituent order in subordinate clauses	130
5.4.1. Relative clauses	131
5.4.2. Interrogative clauses	132
5.4.3. Exclamatory clauses	132
5.4.4. Other finite, non-finite and verbless clauses	133
5.5. Position and order of adverbials	134
5.5.1. Main positions in finite clauses	134
5.5.2. Initial position	134
5.5.3. Medial position	135
5.5.4. Terminal position	137
5.5.5. Overview of positions in finite clauses	138
5.5.6. Positions in non-finite clauses	138
5.5.7. Factors governing the distribution of adverbials	140
5.5.8. Position, form and relative weight	140
5.5.9. Position and (con)textual cohesion	141
5.5.10. Position and scope	142
5.5.11. Position and semantic clarity	144
5.5.12. Position and style	145
5.5.13. Relative position of adverbials	146
5.6. Discontinuity	148
5.6.1. Discontinuous clauses	149
5.6.2. Discontinuous verb groups	155
5.6.3. Discontinuous noun and pronoun groups	155
5.6.4. Discontinuous preposition groups	159
5.6.5. Discontinuous adjective and adverb groups	160
Chapter 6: Coordination and subordination	165
6.1. Introduction	165
6.2. Coordination	167

6.2.1. Coordination and ellipsis	167
6.2.2. What can be coordinated?	169
6.2.3. Types of coordination	171
6.2.4. Coordinating conjunctions	174
6.3. Subordination	178
6.3.1. Introduction	178
6.3.2. Subordination at clause level	178
6.3.3. Subordination at group level	180
6.4. Markers of clausal subordination	183
6.4.1. Subordinating conjunctions	183
6.4.2. Interrogative and relative pronouns	185
6.4.3. Complex subordinating conjunctions	187
6.4.4. Correlative subordinators	188
6.4.5. Other markers of subordination	189
Chapter 7: The simple sentence	191
7.1. Introduction: simple and complex sentences	191
7.2. Referential properties: situations	191
7.2.1. Actionality: dynamic vs. stative situations	191
7.2.2. Subtypes of dynamic situations	193
7.2.3. Subtypes of states	194
7.3. Participant roles	196
7.3.1. Introduction: general roles	196
7.3.2. Specific roles	197
7.3.3. Additional points	199
7.3.4. Formal links	200
7.4. Voice	205
7.4.1. Introduction	205
7.4.2. Functions of the passive	206
7.4.3. Extended use of the passive	208
7.4.4. Voice restrictions	210
7.4.5. Nonfinite passives, GET-passives and notional 'passives'	214
7.4.6. Passives vs. adjectival non-passives	217
7.5. Polarity	219
7.5.1. Introduction	219
7.5.2. Standard negation	220
7.5.3. Domain of negation	221
7.5.4. Syntactic field of negation	223
7.5.5. Negative sentences with global domain and limited field	224
7.5.6. Local negation	227
7.5.7. Nonassertive forms	228

7.5.8. Semantic scope of negation	230
7.5.9. Non-declarative sentences	234
7.5.10. Emphasis and focus	235
7.6. Subject-predicator concord	238
7.6.1. Introduction	238
7.6.2. The basic rule	238
7.6.3. Singular or plural subject realization?	240
7.6.4. Notional concord	245
7.6.5. Attraction	249
7.7. Other types of external concord	250

Chapter 8: The complex sentence 253

8.1. Introduction	253
8.2. Classification of subclauses	254
8.3. Subject clauses	256
8.4. Cleft sentences	258
8.5. Object clauses	263
8.6. Subject complement clauses	265
8.7. Indirect object clauses and object complement clauses	266
8.8. Adverbial clauses	266
8.9. Conditional clauses	269
8.10. Clausally realized disjuncts	271
8.11. Polarity in complex sentences	275

Part III

Chapter 9: Verbals 277

9.1. Introduction	277
9.2. Verb forms	278
9.3. The external relations of verbals	280
9.4. The internal structure of verb groups	281
9.5. Auxiliaries and their delimitation	282
9.6. Tense and aspect	285
9.6.1. Introduction	285
9.6.2. Deictic forms: present and past	287
9.6.3. Future forms	290
9.6.4. Perfect forms	293
9.6.5. Future perfect forms	296
9.6.6. Progressive forms: introduction	298
9.6.7. Present and past progressive forms	302
9.6.8. Future progressive forms	304

9.6.9. Perfect progressive forms	305
9.6.10. Future perfect progressive forms	306
9.6.11. Nonfinite progressive forms	306
9.6.12. Recapitulation	307
9.6.13. Tense-aspect in indirect speech	311
9.6.14. Tense-aspect in literary narrative	314
9.7. Modal uses of tense-aspect forms	316
9.8. Mood	319
9.8.1. Introduction	319
9.8.2. The imperative	320
9.8.3. The subjunctive	322
9.9. Modality	324
9.9.1. Introduction	324
9.9.2. MAY/MIGHT	327
9.9.3. CAN/COULD	329
9.9.4. MUST	332
9.9.5. SHALL/SHOULD	336
9.9.6. WILL/WOULD	339
 Chapter 10: Nominals	 341
10.1. Preliminaries	341
10.1.1. Nouns and noun groups	341
10.1.2. The external relations of nominals	341
10.1.3. The internal structure of noun groups	342
10.1.4. The functional domain of nominals	344
10.2. Categorization	350
10.2.1. What's in a head?	350
10.2.2. The semantics of nouns	352
10.2.3. Gender	353
10.2.4. Types of nouns and referents	355
10.2.5. Countability	356
10.2.6. Recapitulation	359
10.3. Determination	360
10.3.1. Types of determiner	360
10.3.2. Co-occurring determiners: pre- and post-determiners	362
10.3.3. Determiners and quantifiers	363
10.3.4. Referential orientation	365
10.3.5. The definite article	367
10.3.6. The indefinite article	372
10.3.7. Zero determination	378
10.3.8. The genitive	382

10.4. Quantification: the number category	390
10.4.1. The regular singular/plural distinction	390
10.4.2. Irregular plurals	391
10.4.3. Number-invariable nouns	394
10.4.4. Collective nouns	395
10.4.5. What is pluralized?	396
10.4.6. The uses of the singular and the plural	397

Chapter 11: Pronominals 399

11.1. Preliminaries	399
11.1.1. Definition of the pronoun group	399
11.1.2. Classification of pronouns	400
11.1.3. The external relations of pronominals	401
11.1.4. The functional domain of pronominals	402
11.2. Central pronouns	404
11.2.1. Personal pronouns	404
11.2.2. Possessive pronouns	412
11.2.3. Reflexive pronouns	415
11.3. Pronouns without a person distinction	418
11.3.1. Demonstrative pronouns	418
11.3.2. Interrogative pronouns	424
11.3.3. Relative pronouns	428
11.3.4. Indefinite pronouns	436

Chapter 12: Adjectivals and adverbals 447

12.1. Preliminary discussion of adjectivals	447
12.1.1. Adjectives and adjective groups	447
12.1.2. Semantics	449
12.1.3. Morphology	451
12.1.4. The external relations of adjectivals	452
12.1.5. Parataxis and hypotaxis	455
12.1.6. Descriptive, classifying and specifying adjectives	456
12.1.7. Positional restrictions	458
12.2. Adjectival modification and positional ordering	462
12.2.1. The functional domain of adjectivals	462
12.2.2. Modificational zones	463
12.2.3. Inherent Mod. I, Mod. II and Mod. III adjectives	465
12.2.4. Structure in and across Mod. zones	466
12.2.5. Zone-internal order	469
12.3. Comparison of adjectives	473

12.3.1. The basic system of comparison	473
12.3.2. Spelling and pronunciation	475
12.3.3. Irregular comparison	476
12.3.4. The choice between morphological and syntactic comparison	478
12.3.5. The use of compared forms	480
12.4. The substantival use of adjectives	483
12.4.1. What is meant by 'substantival use'?	483
12.4.2. Generic and specific reference	484
12.5. Adverbals	486
12.5.1. Preliminaries	486
12.5.2. Morphology	488
12.5.3. VERY versus MUCH	491
12.5.4. The external relations of adverbals	492
12.5.5. Comparison	494
References	497
Subject index	503
Word index	525

PART I

1. Preliminaries

1.1. Goals and framework

This grammar is aimed at native and non-native students of English in university and other tertiary education. At the same time, however, it is sufficiently extensive, thorough and detailed to serve as a source of reference for professional linguists and teachers of English. In order that it can be used not only as an advanced *textbook* but also as a *reference book*, specific topics are as far as possible dealt with separately and exhaustively. In this respect the present book differs from a number of other recent grammars of English.

The grammar itself is not explicitly contrastive and therefore not limited to a specific group of non-native students of English. It does, however, pay special attention to characteristic features of English which are more acutely felt by non-native than by native speakers of English, and in this sense it is implicitly contrastive. For example, more attention is given to constituent order – an important problem area for virtually all non-native speakers of English – than in most other grammars of English of comparable size.

The present book is not written within the framework of any particular linguistic theory (e.g. Functional Grammar or Chomskyan Generative Grammar). Such a theoretical attachment would isolate us from too many readers. In order to achieve the goals stated above we have been largely *eclectic* in our descriptive approach, though not in a random fashion. Our guiding principle will be a strict *form/function distinction* as it applies at all levels to the constituents of the sentence. In this respect the syntactic framework adopted here is largely the same as that presented in Bache, Davenport, Dienhart and Larsen 1993. The structure of the present grammar is thus determined by considerations of how to describe forms and functions most appropriately at all constituent levels in English.

1.2. Organization

The description of English grammar given in this book is divided into three parts. In *Part I* we offer an introduction to syntax (chapter 2), present the descriptive framework, i.e. the form/function distinction at all levels of analysis from sentence to word (chapter 3), and develop the sentence analysis system to cope with complex syntactic issues, such as stacking, ellipsis, zero

constituents, complex predicates and the relationship between sentence type and pragmatic utterance function (chapter 4).

Part II describes basic syntactic characteristics in English: constituent order (including inversion, discontinuity and the position and order of optional adverbials) in chapter 5; coordination and subordination (including a discussion of determination, complementation and modification) in chapter 6; the simple sentence (including a discussion of situation types and participant roles, voice, negation and concord) in chapter 7; and the complex sentence (including a formal and functional classification of subordinate clauses) in chapter 8.

Part III is devoted to group structure and word classes. It discusses verbs and verb groups (chapter 9), nouns and noun groups (chapter 10), pronouns and pronoun groups (chapter 11), adjectives/adverbs and adjective/adverb groups (chapter 12). In this part of the book categories such as gender, number, tense, aspect, mood and comparison are investigated.

Though the book is structured according to *form*, it is characterized by a strong element of *function*. A great deal of attention will be given to syntactic functions such as subject, object, predicator, etc. and head-dependent relationships, to semantic functions such as agent, affected, instrument, etc. and to pragmatic functions such as topic and comment. However, our book is not a grammar of functions in the sense that it selects as its point of departure a number of major communicative functions such as 'referring to people and things', 'giving information about people and things', 'expressing time' and 'expressing manner and place' (cf. e.g. *Collins Cobuild English Grammar* from 1990 and Downing & Locke 1992).

Consideration will also be given to *information structure*, and in this connection the role played by stress and intonation and by textual factors outside the sentence influencing its syntax will be taken into account. As a very important function of the ordering of sentence constituents is to signal the way in which a message is organized into information units, information structure plays an important role in our account of constituent order in chapter 5 and of voice in section 7.4.

1.3. Data

The approach taken to data in this book is non-positivist and instrumental. We thus regard data as a means rather than an end: a means to secure analytical breadth and precision as well as illustrative exemplification. Our approach to grammar is not corpus-driven, and we do not see it as our task to provide an exhaustive description of one or more corpora. Our examples are derived from a number of sources: from modern British and American

written texts (newspapers, magazines, fiction, etc.), from other grammars of English (including descriptions based on corpora of spoken and written English), from dictionaries and from introspection. Some of our examples are thus not 'real'. In those cases where we have invented examples, we have used our own intuitions about acceptability. When in doubt, however, we have conferred with native speakers of English. Sometimes we have also found it useful to modify authentic examples, thereby producing semi-authentic examples. In those cases where our examples are from dictionaries or from other books on English grammar we have indicated their source. In the remaining cases we have not regarded the benefit and interest to the reader of being informed about the source of each individual example as sufficiently great to necessitate an indication of sources (in the text and in the bibliography), which is both time- and space-consuming. Perceptive readers will recognize the following authors as being among our favourite literary sources: Martin Amis, Julian Barnes, A. S. Byatt, Len Deighton, John Irving, P. D. James, John le Carré, Timothy Mo and Isaac Singer.

The view held on data in this book is that the examples should be the ones that are relevant to a description of English grammar in all its aspects and that it does not matter whether they are derived from corpora, authentic texts, elicitation or introspection as long as they are relevant. Although corpora constitute a highly valuable source of material, and although access to corpora has been shown to reveal aspects of English grammar which have not been captured before, we do not think that a description of English grammar should proceed from authentic examples exclusively, for in that case relevant data may be excluded.

1.4. Varieties of English

All languages are characterized by variation. Some varieties are *user-related* and associated with language users living in a particular region or belonging to a particular class. But besides such regional and social dialects there are also varieties which are associated with special functions and which are *use-related*. Such variation is to do with field of discourse (e.g. law, business, science) or communicative situation (e.g. formal or informal) and differs from regional and social variation in being transient. Under use-related variation we can also include the difference between spoken and written language. There are other types of variation as well, for example according to sex and age, but the major types of variation may be said to be determined by region, social group, style (function, situation/participants) and medium (spoken/written).

English is the most widely used language in the world. It is used by at least 750 million people in addition to being the mother tongue of about 350 million people. In countries like India, Nigeria, Kenya and Singapore it is a *second language*, and is used for administration, education and broadcasting. It is therefore hardly surprising that it is characterized by a great deal of variation. Today its regional varieties differ from each other primarily with respect to pronunciation and vocabulary. While Australian English, for example, can be identified by a collection of pronunciation features (one of which concerns the pronunciation of the diphthong in words like *Australia* and *mate* as /æɪ/) and specific words and word meanings (for example *red-back*, a particular kind of spider, and *scrub* 'poor vegetation'), its grammar is remarkably similar to that of other regional varieties, particularly British English. Variation according to field of discourse primarily is to do with vocabulary. Legal English, for example, makes use of special legal terms and archaic expressions such as *aforesaid*, *aforementioned* and *hereinafter*. On the other hand its syntax, though tending to be rather complex, does not differ significantly from that of other varieties. Predominantly social varieties such as Cockney (the English used by working-class Londoners) and Black English (used by some US citizens of African background) have many special features of pronunciation and also many special words. But here there are also several grammatical features which are not shared by other varieties of English.

1.5. Standard English

The term 'Standard English' is widely used but is by no means easy to define. While it may be difficult to speak of a standard language as a whole, it is much less problematic to characterize some specific features as standard (e.g. the ending -s in the 3rd person singular form of English verbs: *she wants*) and other specific features as non-standard (e.g. the absence of the -s ending in such a form). A large number of those features which are considered standard in English today are derived from the Middle English dialect spoken in the East Midlands which became predominant as the official form of the language and was therefore the one preferred in writing and printing.

'Standard English' is used to describe the variety which is today most widely accepted and understood either within an English-speaking country (for example Standard American English) or throughout the English-speaking world (Standard General English, understood as a supra-regional language, or 'standard of standards'). Linguists tend to agree that Standard English is most easily identified in print (irrespective of pronunciation, which varies considerably from place to place), that it is the variety used by

most newsreaders on radio and television networks (BBC, CBS, NBC, ABC, CBC, etc.) and that it relates to social class and level of education (see McArthur 1992, on which sections 1.4, 1.5 and 1.6 of this chapter are largely based). It is the written form used by all educated British writers in neutral or formal style. As Standard British English is remarkably similar to that of other national standards, for example the American, Australian and Canadian standards, it has been claimed to be the written form used by writers of English throughout the world. It is the English we find, for example, in the *New York Times*, the *Independent*, the *Toronto Globe and Mail* and the *Sydney Morning Herald* and which is described in *Webster's Third New International Dictionary* and the *Oxford English Dictionary*. This 'monocentric' view, according to which English has a (British) core and a periphery, has been challenged by some scholars, however, who hold that English has become 'pluricentric' and that it is therefore more correct to speak of 'Englishes' than of 'English'.

In this book we describe the grammar of Standard British English (BrE). Owing to the grammatical similarity between the different national standards this description will apply very largely to other national standards as well. As American English (AmE) is particularly important with respect to range and number of speakers, we shall, however, account for specific differences between BrE and AmE in the course of our description. We shall also pay attention to the difference between spoken and written English and describe grammatical features which are characteristic of spoken English.

1.6. Grammatical variation

The most important regional varieties of English are American, Australian, British, Canadian, Caribbean, Indian, Irish, New Zealand, Scottish and South African English. Regional subvarieties occur as well, as exemplified by Northern British English and by the American dialects spoken in Eastern New England and the South. Social subvarieties of a regional variety can be illustrated by the occurrence in Indian English of three levels: the acrolect (educated Indian English), the basilect (pidginized varieties) and the intermediate mesolect. Among the predominantly social varieties Cockney, Black English and Chicano English (the English used by speakers of Mexican heritage in the US Southwest) are particularly noteworthy. As pointed out in section 1.5, the grammatical differences between these regional and social varieties will not be systematically accounted for in this book. What we shall do instead here is to illustrate grammatical variation by looking briefly at a few selected grammatical features (all of which will be dealt with more extensively in subsequent chapters).

In BrE the *present perfect* (e.g. *has signed*) and the *simple past* (e.g. *signed*) are both used to describe events that precede the moment of speaking, but they differ in perspective. The former presents a past time event as having implications about what is true of the moment of speech, e.g. *She has signed the letter*. The latter presents a past event as something which has no such implications and which is in this sense over and done with, e.g. *She signed the letter*. The two verb forms are used in basically the same way in e.g. Australian, New Zealand, Scottish and South African English as in BrE. In AmE, however, the simple past is often used where the present perfect is used in other varieties, for example – addressing a child about to go to bed – in *Did you brush your teeth?*, where a very recent past event is referred to. Other examples illustrating this American use of the past are *Did the children come home yet?* and *You already told me*. In Indian English (the mesolect), on the other hand, the present perfect is often used under conditions where the simple past is used in other varieties of English, for example in a sentence like *He has bought the car yesterday*.

The so-called *progressive* construction – which consists of a form of *be* followed by a present participle, as in *She was reading* – is used to present an event as an ongoing process. In most varieties of English this construction is ruled out if the verb is meant to be 'stative', i.e. meant to express a relational state of affairs (e.g. *contain*) or inactive perception or cognition (e.g. *hear*, *know*). In some varieties, however, there are fewer restrictions on what verbs can be used in the progressive. In Scottish English we find progressive examples with stative verbs like *He was thinking he'd get paid twice* and in Indian English (the mesolect) examples like *Lily is having two books* (these examples and the ones given below in this section are quoted from McArthur 1992). In other varieties of English, progressive meaning is reinforced by a special word: in Black English this is *steady*, as illustrated by *We be steady rappin'*, and in South African English it is *busy*, as in *We were busy waiting for him*. In Black English progressive meaning can also be expressed by *steady* exclusively, as in *They be high steady*.

Thirdly, some varieties differ from others with respect to *modal verbs*. In AmE, *shall* and *ought* are rarely used outside formal style. In Scottish English there are several nonconformist uses of the modals: *shall* and *may* tend not to be used in informal speech, *must* is not used to express compulsion (for which *have to* or *have got to* are used) and *need* and *dare* are not used as auxiliaries but as main verbs exclusively, as in *He didn't need to do that* (not *He needn't do that*) and *She doesn't dare to talk back* (not *She daren't talk back*). This last property is shared by Black English as well.

As appears from these examples, many varieties of English stand out in that they differ grammatically from Standard British English.

1.7. Variation according to medium

There is in many languages a good deal of variation according to *medium*: the grammar of the *written* variety often differs significantly from that of the *spoken* variety. The extent to which speech and writing differ varies from language to language: in e.g. Arabic there are different standardized varieties for the two media and the relationship between them is tenuous (a situation referred to as 'diglossia'), while in English speech and writing are felt to be predominantly stylistic variants.

Before describing some of the characteristics of these variants, it is important to make a distinction between *medium* and *channel of communication* (cf. Lyons 1981: 18). Speech and writing are best understood as different media serving different communicative purposes in different contexts, as distinct from the actual (oral or written) channel of communication. Thus speech is used in e.g. everyday conversation while writing is used in e.g. newspapers. Usually writing is delivered 'in writing' and speech is actually spoken. But this is not necessarily so: we can speak the way we write (e.g. when we read aloud), and we can write, more or less, the way we speak (e.g. in e-mail and in dialogues in a novel). Despite such cases where we mix medium and channel of communication, the fact remains that speech and writing are often structurally and functionally different varieties of language, each facilitated (and restricted) by its usual channel of communication. Thus, for example, in speech we may rely heavily on *prosody* (intonation and stress) and *para-linguistic* means (such as e.g. voice quality, gestures, eye-contact, smiles, frowns, yawns, etc.) to get our message across. In writing, we are left with a number of conventional symbols for organizing our message: full stops, commas, question marks, exclamation marks, bold face, capitals, etc., not to mention the choice between handwriting and typing/printing. On the other hand, speech is more transient (unless, of course, someone makes a specific point of recording it), and it is more difficult to edit than written language, which is usually the product of relatively careful planning and drafting, and which always leaves you with a record of the communicative event. The communicative interaction between the participants of a speech situation is more immediate and complex than when writing is involved.

In terms of more specifically grammatical features, spoken and written English differ in a number of important ways. Characteristically, writing, unlike speech, is lexically very economical and, at the same time, dense: written sentences generally contain fewer (partially or wholly) 'redundant' words but more 'heavily loaded' words than spoken utterances. A feature that adds to the density of writing is the greater complexity of some of the units that make up the sentence, especially the so-called noun groups (cf. section

3.3.1), which tend to be longer and to contain more levels of structure than in speech. Conversely, speech has greater density or complexity in the organization of clauses (in terms of what is usually referred to as 'coordination' and 'subordination', cf. sections 3.3.3 to 3.3.5). These differences are borne out very nicely by examples like the following (from Halliday 1985) where the a-variants are typical instances of written language and the b-variants of spoken language:

- (1a) Investment in a rail facility implies a long-term commitment.
- (1b) If you invest in a rail facility, this implies that you are going to be committed for a long term.
- (2a) The growth of attachment between infant and mother signals the first step in the development of a child's capacity to discriminate amongst people.
- (2b) When an infant and its mother start to grow attached to each other, this is a sign that the child is beginning to discriminate amongst people.

Note also that the a-examples of writing are fairly *static* in their presentation of the message, whereas the b-examples of speech are much more *dynamic*. When information is crammed into a few complex units (as in (2a)), what we refer to becomes rather fixed, factual and unchanging. But when basically the same information is spread over a number of clauses (as in (2b)), we get a clearer sense of the activities and processes involved: there are more verbs and consequently a clearer time sequence emerges.

Apart from the very general differences between speech and writing mentioned so far, there are numerous more specific differences pertaining to the grammar of English. Let us mention a few of these. In *written* English, an adverbial (i.e. a sentence function which is not a subject, predicator, object or complement, see section 3.2.9) is frequently realized as an *-ing* participle clause or an *-ed* participle clause:

- (3) *Giving him a light*, I set fire to his moustache.
- (4) *Her oration finished*, she breathed heavily with an overflow of indignation.

While this type of realization is not an exclusive property of the written medium, it is found less frequently in spoken than in written English.

Secondly, the so-called *subjunctive* is typical of written English:

- (5) Whatever *be* the reason, we cannot tolerate his disloyalty.
- (6) Grafton would have rung if the plane *weren't* on its way.

In informal spoken BrE, these subjunctive forms would be replaced by *may be* (placed after *reason*) in (5) and by *wasn't* in (6).

In sentences like *We must put some flesh on your bones* and *I just saw a show on television* where there is both an object (*some flesh* and *a show*, respectively) and an adverbial (*on your bones* and *on television*, respectively),

the former typically precedes the latter. Sometimes, however, this ordering may be reversed (as we shall return to in section 5.3.10). This can be illustrated by the following examples in which the adverbial is placed before the object:

- (7) I just saw [on television] [how some Indian people started a shop and put the old grocery on the corner out of business].
- (8) Hello, my name is Penny Rogers. I bought [some time ago] [a PowerBook 180]. I can't get the internal modem to work and would like to have someone look at it.

This ordering is found in both writing and speech, but often for different reasons: in writing it is the result of careful planning and involves consideration of e.g. weight as in example (7), where the object is very long and hence preferred at the end of the sentence to prevent it from unduly delaying the occurrence of the adverbial. In speech the ordering of the adverbial before the object is often the result of lack of planning, or rather, planning on the spur of the moment: the order of the units here reflects the order in which the speaker thinks of what to say rather than any strict grammatical principle.

As a final example of a grammatical feature which is typical of written English, it is the case that in English the verb, or 'predicator', may be placed before the subject if an adverbial is fronted to give prominence to it or to establish *narrative continuity* (see section 5.3.6 on so-called full inversion after a fronted adverbial):

- (9) On the walls were pictures of half-naked women and colourful landscapes.
- (10) On the doorstep sat women nursing their babies and gossiping.

This ordering is virtually only found in written English. In spoken English *there* would be used as a 'provisional subject' in (9), and in (10) the predicator *sat* would be placed after the subject.

In *spoken English* we find utterances of the following kind:

- (11) What a load of rubbish!
- (12) Mind if I smoke?

Such 'elliptical' constructions and 'non-sentences' abound in both speech and writing but often for different reasons. In speech, the dropping of redundant words or constructions is the result of a reliance on the immediate context and part of the easy-going flow of the conversation and the smooth turn-taking of the participants. In writing, such 'telegraphic style' is used to *catch the receiver's attention* (e.g. headlines, road-signs, chapter headings, titles, warnings, neon commercials) or to *arrange the message in a clear, systematic, comprehensible manner* (timetables, recipes, shopping lists, bank statements, television programmes, sports results, etc.). In writing, unlike

speech, catching the receiver's attention cannot be done prosodically or through the use of gesture, etc.

Another characteristic feature of spoken English is the frequent occurrence of so-called 'comment clauses' – i.e. clauses like *you know*, *I take it*, *generally speaking* and *to be honest* which serve to add a parenthetical comment to another clause, as in *I don't think you'll pass, to be honest*. In speech it is especially stereotyped comment clauses like the ones just mentioned that are frequent. Comment clauses which are less idiomatic, and which require more planning, for example *(This is a serious mistake,) he will undoubtedly have realized by now*, are not typical of the spoken medium.

So-called *conditional* utterances are typically marked by *if* or *unless*, as illustrated by *If you do that again I'll strangle you* and *Unless you shut up I'll strangle you*. But they can also be expressed by an *imperative* construction followed by a statement introduced by *and* or *or*:

(13) Do that again and I'll strangle you.

(14) Shut up or I'll strangle you.

What the speaker expresses in these examples is an intention to inflict injury on the addressee if a certain behaviour continues or unless a certain behaviour is discontinued. This way of signalling a conditional threat is typically restricted to spoken English.

As a final illustration of a grammatical feature which is characteristic of spoken English, we can mention examples of 'dislocation' (cf. section 4.5):

(15) He's an utter nitwit, *that boyfriend of yours*.

(16) *Your brother George*, I've never understood why he didn't resign.

Here the identity of the person referred to is established by a noun group which is either added as an amplifying tag (as in (15)) or prefixed to the sentence (as in (16)).

In closing this section we should point out that differences between spoken and written English like the ones illustrated in this section are also largely characteristic of *formal* vs. *informal English*, whether written or spoken. Such differences thus characterize not only variation according to medium but also variation according to *style*. Many of the grammatical features which are typical of spoken English are found also in informal written English, for example private letters or memos. Conversely, many of the features characterizing written English are found also in so-called edited speech, for example lectures and political speeches, where medium and channel of communication are conventionally mixed.

1.8. English for Special Purposes

When describing variation it is customary to distinguish between Language for General Purposes (LGP) and Language for Special Purposes (LSP). LSP refers to varieties used by practitioners of a profession in their work (see Kragh 1991). As pointed out in section 1.4, variation according to field of discourse (law, business, science, technology, etc.) is use-related in the sense that it involves switching to a variety which the occasion demands. But as a variety of this sort is typically used by practitioners of a special profession, i.e. by specialists who have gone through a professional socialization process which is partly linguistic, it is in fact user-related as well. Legal language is typically used by members of the legal profession, scientific language by scientists, economic language by economists, and so on.

LSP is primarily characterized by its vocabulary, i.e. by special terms employed by a profession, such as *lien*, *liability*, *habeas corpus*, *statutory* and *aforesaid* in legal English. But it is also characterized by features of grammar which are particularly frequent. While some of these are typical of formal (vs. informal) and written (vs. spoken) English as well, others are largely restricted to a special professional variant.

English for Special Purposes (ESP) tends to be rich in *complex noun groups*. This can be illustrated by groups like *the FT-SE 100 Index of shares in Britain's leading companies* (business) and *the issues of breach of statutory duty and common law negligence in respect of the council's exercise of its power under the Act* (law). When realizing an object, such complex noun groups will often necessitate the reversal with an adverbial mentioned in section 1.7. In the following example the adverbial *on local authorities* in the clause beginning with *to* has been moved forward because of the length of the object noun group (the part of the sentence stretching from *any* to *statute*):

- (1) There was a considerable reluctance on the part of the courts to impose [on local authorities] [any liability for breach of statutory duty other than that expressly imposed in the statute].

The complexity of noun groups is often due to a string of words occurring before the noun constituting the nucleus, or head, of the construction. In its frequent use of such *heavy premodification* in noun groups ESP differs from English for General Purposes (EGP). In scientific English we find noun groups like *inertial confinement fusion*, *near-zero explosive yields* and *the first full digital image model of Mars*, in which the head nouns are *fusion*, *yields* and *model*. In business English heavy premodification can be illustrated by *global gross domestic product*, *International Business Machines' year-end results* and *purchasing power parity exchange rates*,

where the head nouns are *product*, *results* and *rates*. In legal English we find heavily premodified nouns too, for example in *the Nurseries and Child Minders Regulation Act 1948*. Owing to its general condensation and the way in which its noun groups tend to be packed with information, LSP is sometimes informally referred to as 'agglomerese'.

As LSP is typically used to describe and direct, whereas emotive and social uses are not normally involved, it must aim at being clear, concise, objective and reliable. Such pragmatic requirements affect grammatical choices. Descriptive and directive technical texts, for example, have been shown to contain many passive constructions where the preferred verb form is in the present, and many compound nouns and adjectives which have been derived from clauses (see Munck 1991). In English this can be illustrated by an example like the following (quoted from McArthur 1992: 1026):

- (2) Three modes of operation are required: voice-activated mode (VOX), press-to-talk (PTT) and call.

Here the choice of the passive contributes to making the message objective (impersonal), its present form to making it general (what is described is valid at all times), and the use of compounds to making it concise. Conversion of clauses into compound words for the sake of brevity can be further illustrated by examples like *quick-drying (ink)*, *quick-action (reversing gear)*, *rapid-hardening (cement)*, *diesel-powered (engine)* and *self-raising (flour)*. It can also be noted that in order to avoid ambiguity, the second of two noun groups referring to the same entity is sometimes not replaced by a pronoun as it typically is in EGP.

In legal English the modal verb *shall* is used with third-person subjects to denote what is legally mandatory:

- (3) The tenant *shall* quietly possess and enjoy the premises during the tenancy without any interference from the landlord.

This usage is not found in current EGP, nor in other types of ESP. While there are thus features of grammar which are restricted to (a variety of) ESP, the grammatical differences between ESP and EGP are nearly always quantitative rather than qualitative. We find the same features of grammar in both varieties, but the frequency with which they occur is often markedly different. As we saw in section 1.7, this is also largely the case with varieties engendered by differences in medium.

2. An introduction to syntax

2.1. The word

As native speakers of a language, and very often also as learners of a foreign language, we have an intuitive knowledge of that language, including its syntax and the basic units of its grammar. Thus, for a start, we all have a pretty good idea of what a *word* is. To realize this, we only have to consider the following passage, where we have eliminated any indication of word boundaries (such as, typically, an empty space between words):

- (1) thepolicemanlookedatthembothhesniffedthatwastheuncooperativeattitudyou mightexpectfromafamilythatencouragedtheirdaughteretogaroundwithyanksan dthese werewelltodopeoplenotworkingclasssuchlaxattitudesoffendedhimhe'd makesurethatnodaughterofhiskeptcompanywithforeignsoldiers

Once we recover from the initial confusion of having to decipher such a muddled, uninviting passage, we can all find the individual words of the original text:

- (2) The policeman looked at them both. He sniffed. That was the uncooperative attitude you might expect from a family that encouraged their daughter to go around with Yanks. And these were well-to-do people, not working class. Such lax attitudes offended him. He'd make sure that no daughter of his kept company with foreign soldiers.

There are, admittedly, occasional problems: is *well-to-do* one or three words? Is *He'd* one or two words? But apart from such nitpicking, we are perfectly capable of identifying the words of any language familiar to us. In writing, word boundaries are signalled by blanks or by punctuation marks. In speech, they are often, though not always, signalled by factors such as the exact onset of stress (as in *see the 'meat* vs. *see them 'eat*) and/or the variant of speech sound selected (as in *keeps ticking* vs. *keep sticking*, where the /t/ is aspirated (i.e. pronounced with a puff of air) when it is a word-initial sound as in the former case). We all know how to signal and interpret word boundaries in both writing and speech, if only intuitively. And yet, amazingly, it is very difficult to define what a word is.

Obviously, *meaning* is somehow involved: *policeman* means one thing, *family* another. But what is the meaning of *the* and *of*? Clearly, these words mean something but their meaning is not as immediately transparent as the meaning of *policeman* and *family*, which express relatively concrete entities (more specifically, persons). And why is it that *the* and *policeman* are two words in English but only one word in, for example, Danish (namely

politimanden, where the ending *-en* corresponds to the English *the*)? Such frivolity is not reserved for Danish and other foreign languages but is a regular feature of English, too, as we see in the word *uncooperative*, where *un* is only part of a word despite the fact that it has an independently identifiable meaning. Another example is, once again, *policeman*: why is *policeman* one word but both *police force* and *police constable* two? Similarly, why are there two separate words in *class struggle* but only one in *classroom*, which is normally identified as a *compound* word (i.e. a unit of elements which function independently elsewhere)? We are forced to conclude that words cannot be defined simply as 'units of meaning'.

In the language user's conception of words, *convention* seems to be an all-important factor. This, however, should not prevent us from trying to describe the words in English (for which the technical term *lexicon* is often used) with reference to any regular pattern applying to them. As a first step towards such a description, grammarians refer to the smallest meaningful units of language as *morphemes* whether or not they are independent words. In this sense *un-* in *uncooperative*, *police-* and *-man* in *policeman*, *work-* and *-ing* in *working*, and even *-s* in *attitudes* and *-ed* in *encouraged* are morphemes. *The*, *him*, *of*, *to*, *that* etc. are both morphemes and words. They are *free* morphemes in contrast to *un-*, *-s*, *-ing*, *-ed*, etc., which are *bound* morphemes. This means that a word consists of one or more morphemes. The precise identification of words is then to a large extent a question of conventional rules of *morphology*, i.e. rules describing the structure of words in terms of morphemes. There is little consistency across languages in the morphology of words: as we have seen, the meaning of definiteness is in English typically expressed by an independent word, the definite article *the*, whereas in Danish it is typically expressed by a word-internal bound morpheme. Sometimes principles seem to vary even within one and the same language (as in the case of English *policeman* vs. *police force*). Note also that while definiteness in English is expressed by means of an independent word, meanings pertaining to, say, number (singular or plural) and tense (present or past) are fully *grammaticalized* in that they are expressed by word-internal bound morphemes, more specifically by *inflections*. Our intuitive knowledge of the words of a language includes the knowledge of what is conventionally expressed by means of individual words and what is conventionally grammaticalized at the morphological level.

One important characteristic of words is that they are *basic syntactic units*, i.e. the building blocks of larger language constructions, and thus have a high degree of *stability* and *cohesion*. Words are stable in the sense that – unlike many higher-level syntactic constructions such as the sentence – they do not allow rearrangement of their constituent parts. Nor do they allow internal

separation. For example, as language users we are not free to organize the morphemes in words as we please. We have to say *childishness*, not **nessishchild*, **ishchildness*, etc. With higher-level constructions there is often a certain variability: we can say either *the sickening unresolvable mess* or *the unresolvable sickening mess* (with little or no difference of meaning) and we can say both *Bob kissed Gina* and *Gina kissed Bob* (although here there is a marked difference of meaning). Note also that normally words are internally inseparable. Thus while we are often free to separate independent words like *the* and *policeman* in a construction like *the policeman* by inserting an adjective, as in *the young policeman*, we cannot separate the individual parts of *the* or *policeman* and still retain their status as single words. In speech, words are also coherent in the sense that we can insert pauses (*uh*, *uhm*, etc.) *between* words but not usually *within* words (cf. Bolinger 1975: 119). Thus in an unsure and hesitant manner we might say:

(3a) The *uhm* policeman *uh* got *uhm* confused.

But we are unlikely to say:

(3b) The po-*uh*-liceman got con-*uh*-fused.

Stability and cohesion may be important clues in the identification of word boundaries. But they tell us little about what a word really is or about why the principles of word formation differ between languages and even within languages.

The interesting fact is that despite the problem of formulating a water-tight definition of the word, we all have an intuitive knowledge of what a word is. That knowledge comprises in part an awareness of morphemes as units of meaning, in part the recognition of largely conventional rules of how morphemes combine to make up the units that we know as words.

2.2. The sentence

Grammar is not just the study of words and their morphological structure but also of how words combine to make up larger units, such as *sentences*. Like words, sentences are notoriously difficult to define rigidly and objectively. And yet we all have an intuitive knowledge of what a sentence is (cf. Bolinger 1975: 156). To appreciate this, we only have to look at a passage where we have left out all the conventional markers of sentence boundaries (such as punctuation and capitalization of initial letters after full stops):

- (1) Victoria shuddered once again she realized that her father was trying to protect her and she loved him for it and if she admitted to knowing Vince Madigan the next question must inevitably be and what was this American's

relationship with Mrs Hardcastle and then more questions I don't recognize him she said softly

Although this passage is a fairly complicated text with both internal and external dialogue, it is easy to guess at its division into sentences:

- (2) Victoria shuddered. Once again she realized that her father was trying to protect her and she loved him for it. And if she admitted to knowing Vince Madigan, the next question must inevitably be, 'And what was this American's relationship with Mrs Hardcastle?' And then more questions. 'I don't recognize him,' she said softly.

Many people will even discover that there are alternative ways of dividing this text into sentences. Thus *Once again* could equally well belong to the first sentence: *Victoria shuddered once again. She realized ...* Similarly, *and she loved him for it* might be a separate sentence: *She realized that her father was trying to protect her. And she loved him for it.* But no-one would suggest that *Once again she* is a separate sentence. Nor would we allow the long sentence *And if she admitted to knowing ... Mrs Hardcastle?* to be broken into two independent sentences *And if she admitted to knowing Vince Madigan* and *The next question must Mrs Hardcastle*, despite the fact that both contain a verb. While the second part could conceivably function as a sentence on its own, the first part is clearly incomplete. It cannot stand alone. According to conventional wisdom, the two parts are *clauses* within the same sentence.

In speech, clause and sentence boundaries are typically signalled and interpreted in terms of *tone groups* ending with a special *intonational contour*: e.g. a rise (to signal, say, the end of a question or the continuation from one clause to another) or a fall (to signal, say, the end of a sentence). To get a sense of such intonational signals, one can try reading out the passage above with the different segmentations proposed.

Though we have to recognize *And then more questions* as an independent unit, many would hesitate to call it a sentence. It is somehow unfinished, lacking a verbal component such as *(And then more questions) would follow*. Similar problems arise with short units of text like:

- (3) No!
(4) After him.
(5) My turn?

Typically such units are complete utterances (cf. Bache *et al.* 1993: 183ff). But are they sentences? Although they are perfectly acceptable in both speech and writing (in writing as a substitute for a spoken utterance), we hesitate to accept them as sentences. It would help considerably if we treated them as somehow short forms of 'proper' sentences like the following:

- (6) I say no!
- (7) I want you to go after him.
- (8) Is it my turn?

But such 'full constructions' are often cumbersome and not entirely natural or appropriate in context and therefore should not be taken as 'more proper' than those in (3) to (5). We have to accept that not all utterances are sentences. Many linguists make a systematic distinction between 'sentence' as a theoretical unit (defined by grammar) and 'utterance' as a physical unit (a matter of speech production), cf. e.g. Lyons 1995: 32ff. On this view some utterances can be analysed in terms of sentences but utterances do not 'consist of' sentences.

Some grammarians have suggested that a sentence is a unit of grammar expressing a 'complete thought' or a 'complete event'. But surely such definitions are too imprecise to be of much use. Is the thought or event expressed in *Victoria shuddered* any more complete than those expressed by examples (3) to (5), which are not full sentences? Or is it more complete than that expressed by the non-sentential clause *And if she admitted to knowing Vince Madigan?* Or than that expressed by *And then more questions?*

What all this amounts to is that although we have intuitions about what a sentence is, and though we are perfectly capable of dividing a text into the appropriate orthographical or intonational units typically reflecting sentence or utterance boundaries, it is by no means obvious how actually to define a sentence. We usually expect a sentence to contain at least a verbal component and some other unit, but as we have seen, there are complications. Sometimes textual units which do not meet this requirement are treated like sentences in terms of punctuation or intonation, or in terms of their independence as acts of communication. Furthermore, there is a complex relationship between clause and sentence. Somehow, clauses are like sentences in normally requiring a verbal component and some other unit. So what is the difference between the two? In a sense the distinction between clause and sentence is very similar to that between morpheme and word. A sentence seems to consist of one or more clauses the way a word consists of one or more morphemes. Sometimes a clause is also a sentence (and thus resembles free morphemes which are independent words), sometimes a clause is simply too incomplete or dependent to serve as a sentence in its own right (and thus resembles bound morphemes, which never occur independently). While it is possible to consider words to be the basic units of syntax, the building blocks of larger units, sentences can be viewed as the maximal autonomous units permitting syntactic analysis. Sentences and their internal arrangement of words are the *domain of syntax*.

2.3. Grammatical structure

Our intuitions about language are not restricted to the mere identification of possible words and sentences but include the organization of words within sentences, i.e. the *grammatical structure* of sentences. Consider the following example:

- (1) John kissed the little old woman who owns that shaggy dog.

We doubt that our readers have ever seen a sentence completely identical to (1). And yet no one has any difficulty in recognizing it as a grammatical construction in English. In other words, there is an appropriate organization of the eleven words in the sentence: they are all used in the right place, at the right time. We know the individual words and their meaning, and somehow we know the kind of relationship they enter into. To appreciate that this organization of words, the *structure* of the sentence, is not random, we only need to change the order of its *constituent* words:

- (2) old the kissed dog shaggy who John woman little that owns.

Although we have exactly the same words here as in (1), (2) is completely ungrammatical. It has become a list of unrelated words.

If we consider the possible ways of arranging the eleven different words of (1) and (2) in a linear sequence, it is in fact a small miracle that – almost without thinking about it – we hit on the *grammatical* sequence in (1). There are, to be exact, 39,916,800 different ways of combining eleven different elements in a sequence ($1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 10 \times 11$). Some of these many alternatives to (1) are of course perfectly grammatical:

- (3) John kissed the old woman who owns that shaggy little dog.
 (4) John kissed the little woman who owns that shaggy old dog.
 (5) John kissed the woman who owns that shaggy little old dog.

Stretching our imagination a little we may even accept sequences like the following:

- (6) The shaggy little dog who owns that old woman kissed John.
 (7) The old woman who owns little John kissed that shaggy dog.
 (8) John owns the old woman who kissed that shaggy little dog.

These sequences are all grammatical (in the sense that the words enter acceptable, recognizable syntactic relationships), but their meaning may differ from our conception of what constitutes the normal state of affairs in the world (Can a dog own a woman? Can a human being own another human being?) and thus challenge us to think of contexts where it would be appropriate to use such sentences.

But even if we allow for a little stretching of our imagination, there are at the very most, maybe about a hundred possible sentences containing the eleven words in (1). There are millions of unacceptable ones. And yet we all have a fairly impressive ability to spot the very few grammatical sentences and reject all the ungrammatical sequences. This ability presupposes an intuitive knowledge of the possible syntactic relationships between words. In other words, we have an intuitive knowledge of grammatical structure.

2.4. Linearity and the principle of proximity

Let us have a closer look at our intuitive knowledge of grammatical structure. As we have already seen, language is necessarily *linear* in the sense that one constituent unit (a speech sound or a letter, a morpheme, a word, a group of words, a clause, a sentence) always follows another. In speech language takes *time*, and in writing it takes up *space*. Grammatical structure is basically a means by which language comes to terms with, and makes the best of, this basic condition.

Strictly from the point of view of linearity alone, we would expect a sequence of elements to be either *random*, with no discernible patterns in the organization of the elements, or *progressively* related, each element receiving its rank according to its position in the list (in terms of, for example, increasing or decreasing 'importance' or 'priority', or according to some convention, such as 'alphabetical order'). In human language we see both these main types of linear organization. A telephone directory is a good, if fairly artificial, example of progressive linearity, and so is counting. But alphabetical and numerical order is also exploited in many natural expressions, such as:

- (1) Gina got many *A's and B's* in her finals.
- (2) You are a nice chap but you will have to watch your *P's and Q's*.
- (3) Jim and Roger came in *first and second*, respectively.
- (4) They arrived in *twos and threes*.
- (5) He bought *ten or twelve* good books.

In a phrase like *Ladies and Gentlemen*, etiquette dictates a certain priority. Progressive linearity is also present in constructions like:

- (6) Gina is a competent, even brilliant, scientist.

which reflects an increase of the intensity with which *Gina* is described.

Random linearity may be present in constructions like the following:

- (7a) *Alex, Stephanie and Roger* went sailing this morning.
- (7b) *Roger, Stephanie and Alex* went sailing this morning.

- (7c) *Stephanie, Alex and Roger* went sailing this morning.
etc.
- (8a) She almost enjoyed the *warm stale sweet* air.
(8b) She almost enjoyed the *warm sweet stale* air.
(8c) She almost enjoyed the *stale sweet warm* air.
etc.

However, the basic randomness of the italicized constructions may be reduced by considerations of rhythm or by contextual factors.

There is a different, more general, derived sense in which linearity is important in the organization of language: since *simultaneity* of expression is excluded, we can predict that, in compensation, elements that somehow 'belong together' will be placed as closely together in the sequence as possible. Thus, in examples (7a-c), *Alex, Roger* and *Stephanie* belong together (in that they all took part in the event expressed by the rest of the sentence, i.e. they all went sailing) but since we cannot express them simultaneously they are instead placed as closely together as possible. The same applies to *sweet, stale* and *warm* in examples (8a-c): they belong together because they perform the same function in the sentence, namely that of describing *air*. As they cannot be expressed simultaneously but are forced into a sequence, they are at least placed closely together. Given the condition of linearity, it is thus in a sense *natural* that words that belong together should be placed together in the sequence of words making up the sentence. In this way we can say that the necessary linearity in the organization of language leads to the principle of *proximity*.

2.5. Constituency

Grammatical structure imposes an organization on the elements of the string which is neither progressive nor random. In doing this, grammatical structure usually exploits the principle of proximity to create groupings of words that belong together. Let us consider the following short version of example (1) in section 2.3:

- (1) John kissed the little old woman.

In this sentence, the word *the* is not in a random position relative to the other words, nor does it receive any rank according to its place in a progression of elements. Rather it is part of a grammatical structure in which it relates more closely to *woman* than to *John, kissed, little* or *old*. At first blush the order of words in this example seems to violate the principle of proximity: *woman* is further away from *the* than *John, kissed, little* and *old*. But on closer examination, it appears that *the, little, old* and *woman* all belong together in a

group (according to the principle of proximity) and as such enter a 'joint' relationship with *kissed* and *John* at a higher level. The sentence describes an instance of kissing (expressed by *kissed*) in which there are two participants: one who performs the kissing (*John*) and one who receives the kiss (*the little old woman*). There is thus an indication that the sentence can be divided into three parts or *constituents*: [John], [kissed] and [the little old woman]. The interpretation of [the little old woman] as a group of words belonging together is supported by the fact that if we want to move one of the words relative to [John] and [kissed] and preserve the meaning of the three individual parts of the sentence, we normally have to move them all:

- (2) The little old woman kissed John.
- (3) *Woman kissed John the little old.

Another interesting feature that suggests that [the little old woman] is an integrated unit is that we can replace it by one word representing the whole group and that we can use it as the answer to a question about the identity of the person John kissed:

- (4) John kissed her.
- (5) - 'Who did John kiss?'
- 'The little old woman.'

The grouping of words together which share a function is often referred to as *constituency*. Structure in language can be described in terms of constituency: complex language units (like the sentence) consist of a number of constituents which, in turn, may consist of lower-level constituents. Language structure is thus multilayered or *hierarchical*.

Despite the strong tendency for proximity in language, this principle may be overridden by other considerations. Compare the following two sentences:

- (6a) Sarah is painting her house.
- (6b) Is Sarah painting her house?

Example (6a) expresses an activity in progress (*is painting*) enacted by someone (*Sarah*) and involving an object (*her house*). It thus seems reasonable to divide the sentence into the following parts: [Sarah], [is painting] and [her house]. That *is* and *painting* form a group seems intuitively right. Nevertheless it is possible to move one of the words without moving the other, as in example (6b), where *is* is moved up in front of *Sarah*. The physical separation of the two words does not in any way impair the sense that they belong together in a group, as a constituent. The 'broken relationship' seems closely related to the communicative difference between the two examples: the first sentence is a *statement*, the second is a *question*.

It thus seems that *communicative function* is a factor which may override the principle of proximity. The term usually applied to a 'broken relationship' in language is *discontinuity*: in the second example, *Is* and *painting* form a discontinuous group to serve a specific communicative purpose.

Syntax deals with the relationship between the units of a sentence, more specifically the various constituency groupings (continuous as well as discontinuous) that the units enter. Like morphology, syntax is part of our intuitive linguistic knowledge.

2.6. Linguistic creativity and ambiguity

Our intuitive knowledge of syntax is not restricted to an ability to *recognize* various word order patterns when we see them: we all know how to use them whenever we engage actively in communication. Thus, as has been emphasized by proponents of a particularly influential school of grammatical thought, Generative Grammar, we all possess the ability to understand and produce *new* sentences, sentences which have never been uttered or written before, simply by using the familiar patterns of syntax and the lexicon, i.e. the words of the language. Some of the examples discussed in the preceding sections are examples of this kind: not many native speakers of English are likely to have come across them before. In this technical sense, language is *creative*: although it contains a finite number of building blocks (the words in the lexicon), the rules for their legitimate combination are such that an infinite number of sentences can be produced. Maximal flexibility in matching expression and meaning is thus ensured.

Another example of the open-endedness of language is the lack of *isomorphism*, i.e. the lack of a one-to-one relationship, between the units of language and the items of the world that we discuss and refer to, using language. One fairly trivial but instructive example of this is the fact that most nouns can be used to refer to more than just one particular item in the 'real world': in appropriate contexts an expression like *the car* can be used about any car, not just one car. Furthermore, *car* is so general in meaning that it appropriately covers a fascinating range of past, present and future vehicles. The units of language can be said to have a *generic* potential.

Yet another aspect of language, related to linguistic creativity, and which involves syntax more directly, is the principled diversity of meaning we sometimes encounter in a single expression. That *ambiguity* is indeed an important factor in language becomes evident when we consider examples like the following (taken from Bache 1985b: 56; Chomsky 1957: 88; Lyons 1968: 249; Schibsbye 1970: 30, Wells 1947: section 3; and others), which have been the object of much attention in the linguistic debate:

- (1) Old men and women are invited to the party.
- (2) Flying planes can be dangerous.
- (3) She wants to marry a Norwegian who is rich.
- (4) He left his wife to deal with the creditors.
- (5) The girl found a book on Main Street.

In some of these sentences, the ambiguity is more obvious than in others. But most of us will eventually recognize the different meanings potentially expressed in these examples.

In *Old men and women are invited to the party*, the expression [Old men and women] refers either to a group of old men and old women or to a group of old men and of women of any age (young and old alike), depending on whether we interpret the adjective *old* as a modifier of *men and women* or of *men* alone.

In *Flying planes can be dangerous*, [*Flying planes*] is either a word-like nominal expression for aeroplanes with primary stress on the first word (like *police force*) or it is a clause-like expression with primary stress on the second word referring to instances of the activity of flying a plane. The ambiguity arises because the normal concord rules are neutralized in *can*: when *Flying planes* is a word-like nominal expression it takes the plural (as in e.g. *Flying planes are dangerous*); when it is a clause-like construction on a par with *to fly a plane*, it takes the singular (as in e.g. *Flying planes is dangerous*).

The example *She wants to marry a Norwegian who is rich* shows that there are sometimes different interpretations of referring expressions: either [a Norwegian who is rich] refers to a particular person (e.g. Knut Flo from Oslo) or it refers to anyone who qualifies as a rich Norwegian, i.e. any member of the class of rich Norwegians.

In *He left his wife to deal with the creditors*, the person referred to by *He* either lets his wife deal with the creditors (i.e. *his wife* is the agent of *to deal*) or he leaves his wife with the purpose of dealing with the creditors himself (i.e. *He* is the agent not only of *left* but also of *to deal*).

Finally, in *The girl found a book on Main Street*, the girl either found a book about Main Street, or it was on Main Street, of all places, that she found a book. Either *on Main Street* is part of a more complex construction *a book on Main Street*, in which it describes the subject matter of the book involved, or it is a more independent construction describing the location where the girl found the book.

Characteristically, as we have seen, the different interpretations of all the examples described above are tied to different syntactic patterns (i.e. alternative relationships between the units involved) or different uses of the

units making up the sentence. The recognition of ambiguity in such cases is thus a sign that we have a fairly advanced, if 'only' intuitive, knowledge of syntax and grammar.

2.7. Competence and performance

In the preceding sections we have established the fact that the speakers of a language have a high degree of linguistic sensitivity and informal knowledge of their own language. In other words, they have what is often referred to as *linguistic competence*. Not only are they capable of identifying grammatical units like words and sentences, they also recognize complex syntactic patterns and attach appropriate meanings to them, as witnessed in cases of ambiguity. Most important of all, they know how to put their intuitive knowledge to use whenever they engage in communication. More technically speaking, they know how to turn their linguistic competence into actual linguistic *performance*. The fact that few speakers of a language are capable of describing their language skills and of defining the relevant units and patterns of language in a rigid, principled manner should not make us underestimate their competence. Language is in this respect similar to activities like walking or riding a bike: we are very competent at doing these things without thinking about how we do them. And most of us would be hard put to describe all the exact movements involved in these activities in a principled, scientific manner.

The intuitive knowledge speakers have of a language comprises much more than a knowledge of its formal properties, i.e. linguistic competence. Intuitively, we know not only how linguistic expressions are structured but also how to *use them appropriately* in different contexts or situations and in relation to our communicative intentions. For example, speakers of English know how to be formal or informal in their verbal interaction with other speakers. They also know how to describe events as located in time, how to elicit information, how to refer to things and persons, and so on. In short, they have a knowledge of how to do things with linguistic structures. The overall intuitive knowledge that speakers have of a language and of how to use it in context is called their *communicative competence*. Though communicative competence is largely to do with language in use, it is still possible to regard it in terms of 'knowledge of language' and keep it clearly apart from performance; for there is obviously a difference between what speakers are capable of doing verbally and what they actually do in a given situation (see Dik 1989: 5).

Once we have accounted for the structural properties of a well-formed sentence, we have to consider in what context it is appropriately uttered.

When describing a language we are therefore concerned not only with syntactically and semantically well-formed sentences but also with the appropriateness of sentences in a given context. For example, the near-equivalent sentences *You must make your payment by May 31st* and *Your payment must be made by May 31st* are both syntactically and semantically well-formed, but in some contexts only the passive sentence is appropriate, in others only the active. To account for a native speaker's choice of one rather than the other, we need the concept of communicative competence, which combines linguistic competence with context.

2.8. Syntagmatic and paradigmatic relations

In this section, we shall draw attention to a specific example of the kind of knowledge that native speakers of a language seem to have, an important aspect of their linguistic and communicative competence, the knowledge of *paradigms* or *choice relations*.

As we have already pointed out, language is of necessity linear but has syntactic structure imposed on it which exploits and overrides the linearity. Consider a sentence like the following:

- (1) Sally teaches grammar.

This sentence has a relatively simple syntactic structure involving the *horizontal* relationship between the constituents [Sally], [teaches] and [grammar]. According to the rules of English grammar, we interpret the sentence as a statement to the effect that Sally is the one who teaches and grammar is the subject taught. From a cross-linguistic, universal point of view, there is no necessary single arrangement of constituents to express this particular piece of information. In other languages, it may well have to be expressed through a different arrangement of the constituents making up the sentence, corresponding to, for example, *Teaches Sally grammar* or *Sally grammar teaches*, which are ungrammatical in English. The kind of horizontal relationship that can be established between the constituents of a sentence is often referred to as *syntagmatic*.

Each of the constituents in the syntagmatic relationship in *Sally teaches grammar* might have been more complex, thus adding to the overall complexity of the sentence:

- (2) The young woman is teaching English grammar.

To go from the first sentence to the second we replace [Sally] by [The young woman], [teaches] by [is teaching] and [grammar] by [English grammar]. The basic structure of the two sentences is the same. Further complexity is of course possible:

- (3) The very beautiful young black American woman that you met at the pub last night could have been teaching advanced English grammar.

In this sentence, there are again three main constituents corresponding to those in the two first examples: [The very beautiful young black American woman that you met at the pub last night], [could have been teaching] and [advanced English grammar]. This means that despite the verbosity of this example, its basic structure is like that of *Sally teaches grammar*: there are three main constituents only.

The structural similarity of the three examples discussed above shows that although language is linear, thus calling for the syntagmatic, horizontal arrangement of the constituents in the sequence that we recognize as a syntactic structure, there is at the same time a *vertical* dimension to language. A sentence is not just a sequence of elements or units which enter some sort of horizontal relationship. Rather, a sentence contains a number of *slots* which may be filled in different ways for different communicative purposes. Thus, at one level, the three examples contain the same number of slots, namely three, but these slots are filled with constructions of different length and complexity. The constructions which are possible in a particular slot (e.g. [teaches], [is teaching] and [could have been teaching]) enter a choice relation: they are all candidates for a particular function at a particular point, and the choice of one excludes the others. The relationship between the possible constructions in a particular slot is often referred to as *paradigmatic*.

The implication of all this is that the linearity of language should be viewed in terms of a sequence of slots, each an important hallmark, at which the language user has a choice of expression. Language is both syntagmatic and paradigmatic.

Sometimes the choice of expression for a given slot is a choice of one lexical item rather than another:

- (1) Sally teaches *grammar*.
 (1') Sally teaches *physics*.

The paradigmatic choice between *grammar* and *physics* in the frame [Sally teaches ____] is a purely lexical choice, with no implication for the other constituents of the sentence, and therefore not terribly interesting from a grammatical point of view. Other paradigmatic choices involve grammar:

- (1a) Sally *teaches* grammar.
 (1b) Sally *is teaching* grammar.
 (1c) Sally *taught* grammar.
 etc.

The choice of verb form in a frame like [Sally ____ grammar], where a number of different forms of the verb *teach* are possible, must be accounted for. A grammar of English must provide answers to questions like 'Why is *teaches* but not *teach* all right in that particular frame?', 'What is the difference between *teaches* and *taught*?' 'Or between *teaches* and *is teaching*?' etc. Such questions concern inflectional morphology and competing syntactic constructions.

It is also important to specify in our grammar what *types of construction* are possible in particular slots. As we have seen, instead of a name in the first slot ([Sally]), we may have a group of words ([The young woman] and [The very beautiful young last night], respectively), but we cannot normally have a clause:

(4) *That Sally is very competent teaches grammar.

In other frames, clauses as well as names and groups of words are perfectly possible in the initial slot of the sentence:

(5a) *Sally* surprised Jack.

(5b) *The young woman* surprised Jack.

(5c) *That Sally is very competent* surprised Jack.

It is important to realize that the two *dimensions of language*, the syntagmatic and the paradigmatic, are closely interrelated. Thus the choice of a particular construction to fill a particular slot may well affect later choices of constructions (and, conversely, the choice of a construction may be made in anticipation of choices one wants to make later on). Consider the following pair of sentences:

(6a) The young *woman* *teaches* physics.

(6b) The young *women* *teach* physics.

Here the choice of the singular noun *woman* in the initial major constituent necessitates the choice of *teaches* rather than *teach* as the second constituent. By comparison, the choice of the plural noun *women* in the initial major constituent leads the speaker to choose *teach* rather than *teaches* later on in the sequence. The term grammar should be interpreted in a broad sense covering both the paradigmatic and syntagmatic dimension.

In conclusion, the intuitive knowledge that native speakers have, their competence, includes knowledge not only of the syntagmatic dimension of language but also of the paradigmatic dimension. As with the other aspects of native speaker intuition dealt with in the preceding sections, it is difficult to describe one's knowledge of language in a precise, appropriate and objective manner. The aim of this grammar is to provide such a description of

English. What we have set out to do is not simply to teach you, our readers, grammar, because in the sense discussed in the preceding sections you know a lot of grammar already. What we want to do is rather to offer the tools, the terminology and the insights necessary for making your knowledge more explicit.

2.9. Recapitulation

In this introduction to syntax we have shown that speakers of a language have a high degree of linguistic competence: they have an intuitive, implicit knowledge of the basic units of grammar and the various relationships these enter into. This linguistic competence includes intuitions about syntax (the principles of linearity and proximity as well as the principle of constituency which arises from and overrides the two other principles) and of grammatically conditioned ambiguity. Part of the linguistic competence of language users is also a knowledge of paradigmatic choice relations in language. Despite this highly developed competence, most speakers are unable to describe their language skills appropriately. They may be able to identify words and sentences but they cannot define these units. They easily recognize grammatical strings of words in contrast to ungrammatical ones, possible paradigmatic choices in contrast to impossible ones, as well as grammatically conditioned ambiguity. And above all, they know how to use language appropriately: they have communicative competence. But again, if prompted, most people would fail to offer an appropriate account of why and how they do these things. Thus, when we speak of 'learning the grammar of a language', it is not simply a question of acquiring new knowledge but also a question of becoming more conscious of something that we know intuitively already. Even to the foreign learner in need of getting 'all the facts of the language' right, the process of learning grammar to some extent involves getting intuitive linguistic and communicative knowledge turned into explicit conscious knowledge.

In order to teach (native as well as non-native) speakers of English the grammar of the language, we need to turn the intuitive linguistic and communicative competence that native speakers of English have into an explicit one. Against this background, the aim of the present grammar is to offer an appropriate descriptive apparatus and to present the relevant rules of competence that native speakers of English employ when they engage in linguistic performance, i.e. in actual communication.

3. Elementary sentence analysis

3.1. The basic form and function approach

As a first step towards establishing an appropriate descriptive apparatus for the grammar, we shall introduce a fairly elementary approach to the description of the constituents making up sentences. The specific aim of this is to provide the grammar with a common framework and terminology.

3.1.1. Form and function

We draw a basic distinction between the *form* and the *function* of constituents, which applies to all levels of description. We have already touched on this distinction. In our discussion of paradigmatic choice relations in section 2.8 above, we noted that different types of construction may fill a particular slot in a sentence. Thus, as we saw, [Sally], [The young woman] and [That Sally is very competent] are all possible choices in the empty slot in the frame [_____ surprised Jack]. Another way of formulating this insight is to say that different forms may assume the same function in a sentence: [Sally], [The young woman] and [That Sally is very competent] are different forms but may perform the same syntactic function relative to *surprised Jack*.

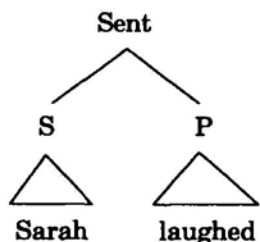
3.1.2. Sentence functions

The main slots for which there is a choice of form in a sentence frame are called *sentence functions*. We recognize five basic sentence functions:

S	=	subject
P	=	predicator
O	=	object
C	=	complement
A	=	adverbial

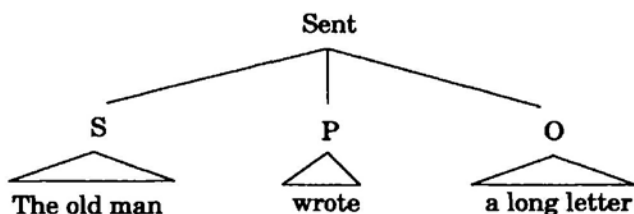
We employ two different techniques in our structural representations: *linear analyses* and *tree diagrams*. The two techniques are notational variants, i.e. different ways of showing the same structure. In a linear analysis (which is convenient for simple or partial analyses in run-on texts), we use square brackets to indicate the beginning and the end of constituents, each bracket tagged with the appropriate label of analysis. In a simple sentence like *Sarah laughed*, we can identify two sentence functions: a subject (*Sarah*) and a predicator (*laughed*). The linear analysis looks like this: S[*Sarah*] P[*laughed*].

The tree diagram, which is a conventional form of syntactic representation, provides an accessible overview of complex analyses. Using the label 'Sent' for sentence, we can draw the following tree diagram for *Sarah laughed*:

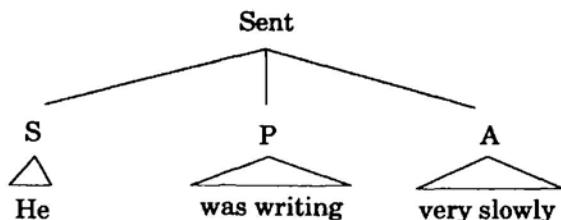


The lines slanting downwards from Sent indicate a 'consist-of relationship': Sent consists of S and P. We use triangles to indicate that our analysis of *Sarah* and *laughed* is incomplete: we have not assigned the appropriate form labels yet. Before we do so, let us first look at some more examples of the five sentence functions:

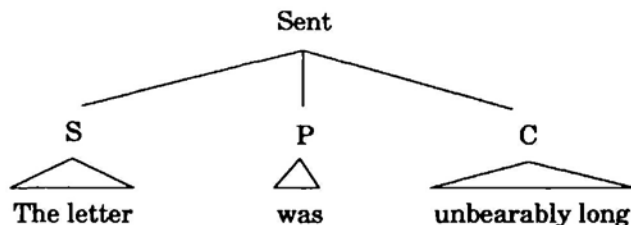
- (1) The old man wrote a long letter.
 S[The old man] P[wrote] O[a long letter]



- (2) He was writing very slowly.
 S[He] P[was writing] A[very slowly]



- (3) The letter was unbearably long.
 S[The letter] P[was] C[unbearably long]



3.1.3. Four form types

There are four different form types capable of assuming sentence functions:

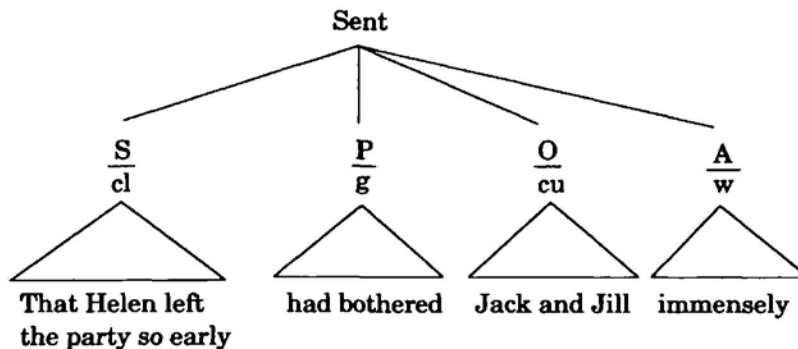
w	=	single word
g	=	group of words
cu	=	compound unit
cl	=	clause

For example, in *Sarah laughed*, both the subject and the predicator are single words: S[Sarah] P[laughed]. In *The old man wrote a long letter*, the subject and the object are groups of words, the predicator a single word: S[The old man] P[wrote] O[a long letter]. All four types of form are present in an example like *That Helen left the party so early had bothered Jack and Jill immensely*: the subject [That Helen left the party so early] is a clause; the predicator [had bothered] is a group; the object [Jack and Jill] is a compound unit with two elements linked together, or coordinated; and, finally, the adverbial [immensely] is a single word.

Notice that we use lower case letters for forms, capital letters being reserved for functions. Separating the two by a colon (:) we have a convention for describing both the function and the form of a constituent:

S:cl	[That Helen left the party so early]
P:g	[had bothered]
O:cu	[Jack and Jill]
A:w	[immensely]

The *colon convention* is used mainly for simple or partial analyses in run-on texts. For more complex structures displayed in tree-diagrams, we use a *function-over-form convention*:



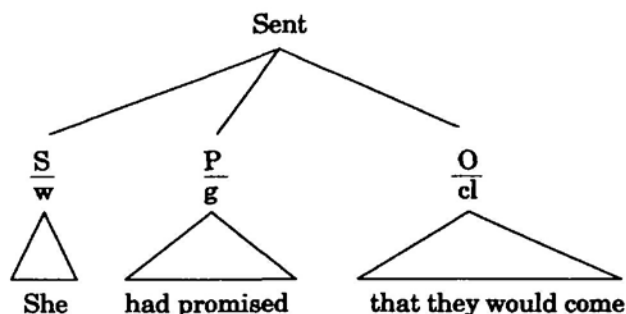
The function-over-form convention is used to indicate that a sentence constituent has a function 'upwards in the tree' in relation to the other constituents of the sentence, while internally it is a construction of a certain

form type to be further analysed 'downwards in the tree' (unless of course it is a single word, permitting no further syntactic analysis).

Here are some more examples:

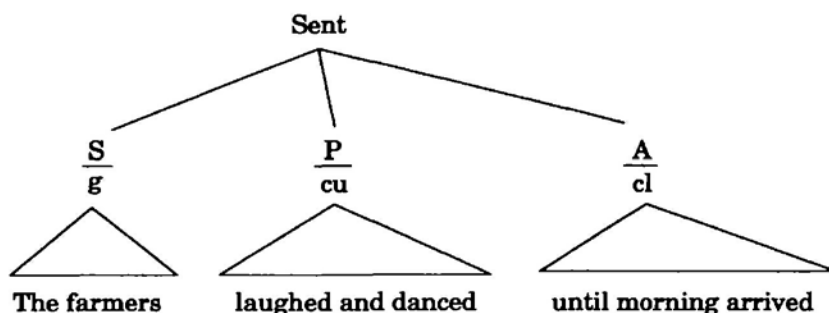
- (1) She had promised that they would come.

S:w[She] P:g[had promised] O:cl[that they would come]



- (2) The farmers laughed and danced until morning arrived.

S:g[The farmers] P:cu[laughed and danced] A:cl[until morning arrived]



3.1.4. Word classes

One of the forms introduced in section 3.1.3 will now be specified further: the individual word (w). Words are traditionally divided into eight main *word classes* according to their notional and formal characteristics:

n	=	nouns	(e.g. <i>car, letter, Jack, idea</i>)
v	=	verbs	(e.g. <i>write, be, receive, hear</i>)
adj	=	adjectives	(e.g. <i>long, old, afraid, big</i>)
adv	=	adverbs	(e.g. <i>slowly, gently, duly, very</i>)
pro	=	pronouns	(e.g. <i>he, she, who, any, this</i>)
prep	=	prepositions	(e.g. <i>by, at, to, from, in</i>)
conj	=	conjunctions	(e.g. <i>that, because, although</i>)
art	=	articles	(<i>the, a, an</i>)

Nouns typically express things or persons. In doing so they are often combined with articles and inflected for the expression of number (e.g. *the car* vs. *the cars*) and the genitive case (e.g. *Jack* vs. *Jack's*).

Verbs typically express actions (e.g. 'writing') or states (e.g. 'being') and inflect for tense and aspect (e.g. *write* vs. *wrote*), person and number (e.g. *write* vs. *writes*).

Adjectives typically express qualities in relation to (pro)nouns (e.g. *a long letter* / *Jack is old*) and often allow comparison (e.g. *longer*, *longest* / *more afraid*, *most afraid*).

Adverbs typically express qualities in relation to verbs (e.g. *Jack moved slowly*), adjectives (e.g. *very big*), other adverbs (e.g. *so gently*), or the rest of the clause (e.g. *Fortunately everybody was saved*). Adverbs are often derived from adjectives by means of the suffix *-ly*: e.g. *slow* → *slowly*, *gentle* → *gently*. Like many adjectives, many adverbs allow comparison (e.g. *more slowly*, *most slowly*).

Pronouns are a rather heterogeneous word class, comprising personal pronouns (*I*, *me*, *you*; *he*, *him*; *she*, *her*; *it*, etc.), possessive pronouns (*my*, *mine*; *your*, *yours*, etc.), reflexive pronouns (*myself*, *yourself*, *herself*, etc.), demonstrative pronouns (*that*, *those*, *this*, *these*), interrogative and relative pronouns (e.g. *who*, *which*, *what*) and indefinite pronouns (*some*, *something*, *any*, *anybody*, *no*, *nothing*, *every*, *everyone*, *all*, (*n*)*either*, *both*, etc.).

Prepositions express relations (often spatial relations) between constituents. They typically do so by relating a noun or group (e.g. *the table*) to another noun or group (e.g. *the book*) as in *the book on the table*, or to some action or state (*The book was placed on the table* / *The book is on the table*).

Conjunctions also express relations between constituents. They do so either by combining constituents at the same level (e.g. *cars and books*, *clever but arrogant*) or by placing one clause (e.g. *He didn't support her*) at a lower level in relation to another clause (e.g. *I said that he didn't support her*).

Articles typically combine with nouns to express definiteness (e.g. *the car*, *the idea*) or indefiniteness (e.g. *a car*, *an idea*).

To the eight main word classes we may add *intj* (interjections like *huh*, *ouch*, *well*, *oh*, *wow*, etc.) and *num* (numerals like *five*, *hundreds*, *1993*, *tenth*, *twenty-first*, etc.). The infinitive marker *to* is special: like many adverbs it is obviously related to verbs; like auxiliary verbs such as *may*, *can*, *will*, etc., it is placed in front of verbs (and thus in fact also resembles the articles, which always precede nouns); like the conjunction *that* it seems void of meaning; and formally it looks like the preposition *to*. We treat it separately, as a word in its own right, and use the abbreviation *infm* to mark it in our analyses.

Each word class will be dealt with more elaborately in later chapters. At this point we shall merely point to certain important facts relating to the

division of words into classes: (i) the identification of word-class membership; (ii) the distinction between words as lexical items and words in use; and (iii) *open word classes* vs. *closed word classes*.

A) The identification of word-class membership. It is often difficult to classify a word in isolation from its linguistic context. Many words are of course easily identifiable as members of one, and only one, word class: *policeman* is always a noun, *eliminate* is always a verb, *the* always an article, *always* always an adverb, etc. But there are also cases where we have to rely on the context to reveal the function of the word before we can classify it. Put differently, there are cases where word-class membership cannot be determined independently of function. For example, *blow* is a noun in *It was a hard blow to him*, but a verb in *The referee may blow his whistle any time now*. *Early* is an adjective in *He took an early train* but an adverb in *He left the party very early*. *Down* is particularly versatile: it is an adverb in *The ship went down*, a preposition in *Sally was walking confidently down the street*, an adjective in *He is in one of his down periods at the moment*, a verb in *He could down a pint of beer in twelve seconds*. It may even be used as a noun in the plural, as in *He has his ups and downs*, or with a completely different meaning, as in *The pillow was full of soft down*.

As can be seen, it is necessary to distinguish, on the one hand, between completely different words with the same form, and, on the other, between different uses of what basically appears to be the same word. Thus it would be sensible to say that *down* with the meaning 'first, soft feathers of young birds' and *down* with a directional meaning are two different but *homonymous* words, i.e. different words which happen to have the same manifestation form. But then, what about the different uses of *down* with directional meaning mentioned above: are they to be considered separate words? It seems most appropriate to recognize the various functional realizations of directional *down* as word-class distinct items (adverb, preposition, adjective, etc.). In practice, then, we treat them as distinct but very closely related words.

B) The distinction between words as lexical items and words in use. Consider now the problem posed by the following examples:

- (1) We all *love* Sally.
- (2) Richard probably *loves* her more than the rest of us.
- (3) Even bad-tempered, old Graham *loved* her once.
- (4) As for myself, I cannot help *loving* her, too.

What we see here is formal (inflectional) variation of an item which does not result in a change of word class. Though formally distinct, *love*, *loves*, *loved*

and *loving* 'belong' to the same word, or lexical item, the verb *love*. This means that we have to distinguish between a word in isolation – the base form as it appears in a dictionary – and its inflectional manifestation form in actual speech or writing. Henceforth we shall use capital letters when we want to emphasize the status of a word as a base form and italics when we want to emphasize the status of a word as a realized manifestation form: *love*, *loves*, *loved* and *loving* are manifestation forms of the base form LOVE. We use this convention in connection with verbs, nouns, adjectives and adverbs.

C) Open word classes vs. closed word classes. Of the eight main word classes listed above, the first four (nouns, verbs, adjectives and adverbs) are open word classes whereas the last four (pronouns, prepositions, conjunctions and articles) are closed classes. Numerals and interjections are open classes, while the infinitive marker is unique and thus does not fit into the open/closed distinction at all.

Open and closed word classes can be distinguished in several different ways. While open word classes have indefinitely many members, closed word classes have relatively few members. While open word classes have a fairly relaxed 'membership policy', admitting new members whenever there is a need for them, closed word classes rarely allow any change. Thus we often get new nouns (for example, as the result of new technology: LASER, VIDEO, SOFTWARE, etc.) but the classes of prepositions and articles stay the same for a very long period of time.

Members of open word classes typically have one or more independently identifiable meanings, and there is no necessary semantic relationship between the meaning of one member of a class and another member of the same class. Thus, simply by looking at nouns like POLICEMAN and STORY we get a clear sense of their meaning. At the same time there seems to be no obvious semantic relationship between them. Members of open word classes are used by the speaker to instruct the hearer to think of things, events, qualities, etc. that the speaker wants to talk about. By contrast, members of closed word classes seem to have little independent meaning: they are grammatical *function words*, assuming their meaning in relation to other words. For example, in isolation it makes little sense to discuss the meaning of, say, the definite article *the*, the conjunction *that*, the relative pronoun *which* and even the preposition *at*. In appropriate linguistic contexts, however, these words assist open-class words in forming coherent sentences and utterances. The presence of e.g. the definite article in the context of a singular noun typically ensures a reading of the noun as a word which refers to a specific, identifiable entity. Unlike open-class words, closed-class words often enter a tight network of functional interdependencies and relationships.

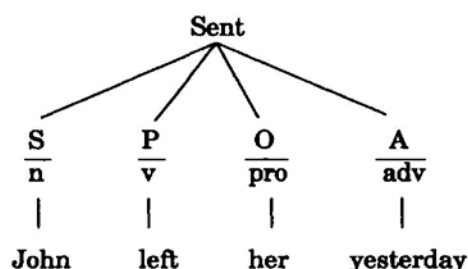
Thus, the function of the definite article is largely complementary to that of the indefinite article: together they share a *functional domain*. The same is true of the other closed word classes, though of course there are more members and therefore more complex networks and systems.

3.1.5. Simple complete analyses

We are now in a position to offer complete analyses of sentences consisting of one-word constituents, like the following (note that in such cases we no longer need the triangle convention):

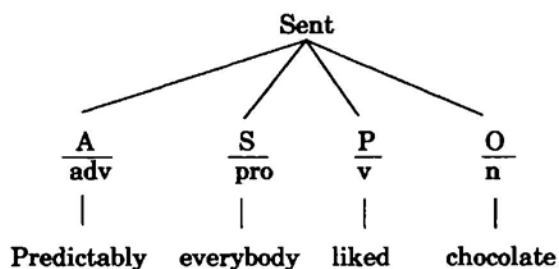
- (1) John left her yesterday.

S:n[John] P:v[left] O:pro[her] A:adv[yesterday]



- (2) Predictably, everybody liked chocolate.

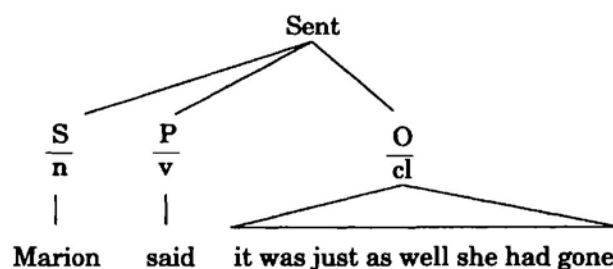
A:adv[Predictably] S:pro[everybody] P:v[liked] O:n[chocolate]



For sentences which contain complex constituents we still use the triangle convention to indicate that further analysis is possible:

- (3) Marion said it was just as well she had gone.

S:n[Marion] P:v[said] O:cl[it was just as well she had gone]

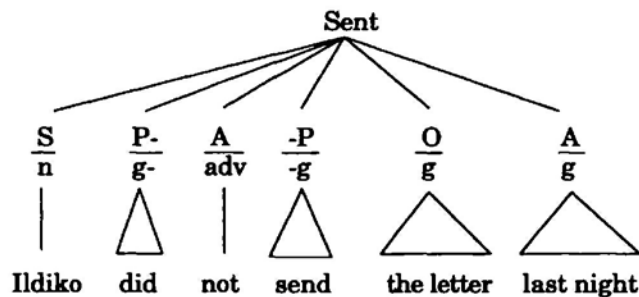


3.1.6. Discontinuity

As pointed out in section 2.4, there is a strong tendency in language for constituents which belong together to be positioned together. However, this principle of proximity is violated under well-defined conditions (see section 5.6 below). In both our linear analyses and our tree diagrams, the resulting discontinuity is marked by hyphens in the following way:

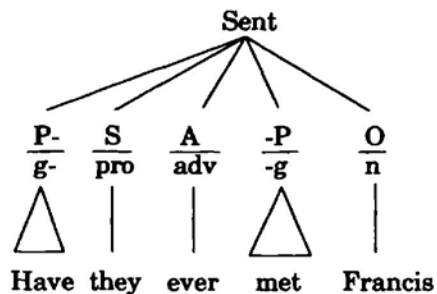
- (1) Ildiko *did* not *send* the letter last night.

S:n[Ildiko] P:g-[did] A:adv[not] -P:g[send] O:g[the letter] A:g[last night]



- (2) Have they ever met Francis?

P:g-[Have] S:pro[they] A:adv[ever] -P:g[met] O:n[Francis]



In these examples, right-hyphenation (i.e. hyphenation *after* a label, such as P:g- in example (1)) indicates a discontinuous relationship between the unit it represents in the tree (*did*) and a unit in the *subsequent* linguistic context (*send*), identically labelled but with left-hyphenation (i.e. a hyphen *before* the label, such as -P:g in example (1)).

Notice that only one hyphen is used for each part of the discontinuous constituent in our linear analyses, representing both discontinuous form and discontinuous function (e.g. 'P:g-' for *Have* in example (2)), whereas in our tree diagrams two hyphens are used, one for the form label (e.g. 'g-') and one for the function label (e.g. 'P-'). It is also important to notice that although each part of a discontinuous constituent may consist only of one word, as in all of the examples above, we have not yet reached word level in our

analysis. The internal relationship of the parts that have been separated remains to be specified, exactly as in continuous constituents.

Having introduced the main sentence functions and the main types of form manifesting them, as well as the convention for marking discontinuity, we now turn to each of the functions S, P, O, A and C.

3.2. Sentence functions and sentence structures

3.2.1. The predicator

The identification of subject, object, complement and adverbial often depends on the prior identification of the predicator. Fortunately the form of the sentence predicator is relatively stable and therefore fairly easy to identify. It always consists of one or more verbs:

- (1a) Jack *treated* Sophia very badly.
- (1b) Jack *is treating* Sophia very badly.
- (1c) Jack *has been treating* Sophia very badly.
- (1d) Jack *may have been treating* Sophia very badly.
- (1e) ?Sophia *may have been being treated* very badly by Jack.

As we see in these examples, there are various ways of expressing a situation of 'Jack treating Sophia very badly': the key word is in each case TREAT. In fact, the italicized predicator in (1a) to (1e) can be regarded as different manifestation forms of the base form TREAT, involving one or more words. To describe the organization of the predicator, we distinguish between *full verbs* and *auxiliary verbs*. A predicator may consist of just a full verb (as in example (1a)) or a full verb as head preceded by up to three (in exceptional cases: four) dependent auxiliary verbs (as in examples (1b) to (1e)).

The difference between a full verb and an auxiliary is normally one of semantic weight: full verbs have independently identifiable lexical meanings whereas auxiliaries have functional characteristics like closed-class items (articles, prepositions, pronouns and conjunctions), relating to and modifying full verbs. Formally, the two types of verb can be distinguished in terms of linear position: in predicators where both are present, the last verb is almost always the full verb and the others are auxiliaries (we disregard cases of inversion like *Also killed in the shootout were three teenagers from the Bronx*, see section 5.3.7). If one wants to test whether a verb is a full verb or an auxiliary, one can convert a statement containing the verb into a 'yes-no question' (i.e. a question of the type which tries to elicit either a yes or a no for an answer). If the verb readily precedes the subject it is an auxiliary whereas if it cannot precede the subject it is a full verb:

- (2a) Rob *was having* a nightmare.
- (2b) *Was* Rob *having* a nightmare?
- (3a) He *can run* a mile in six minutes.
- (3b) *Can* he *run* a mile in six minutes?
- (4a) Steven *finished* his cheeseburger.
- (4b) **Finished* Steven his cheeseburger?
- (5a) Cathy *kept* laughing.
- (5b) **Kept* Cathy laughing?

In examples (2) and (3) BE and CAN are shown to be auxiliaries. In (4) and (5) FINISH and KEPT are shown to be full verbs. To form a *yes-no* question from a statement containing a full verb we have to use *DO-support*:

- (4c) *Did* Steven *finish* his cheeseburger?
- (5c) *Did* Cathy *keep* laughing?

In such cases DO is an auxiliary.

Note that three verbs, BE, HAVE and DO, are special in that they function sometimes as auxiliaries and sometimes as full verbs. In the latter case they may stand alone in the predicator:

- (6a) Jack *is* now fully awake.
- (7a) The old dancer *has* fond memories of Paris.
- (8a) Her parents *did* nothing to change her mind.

The three verbs form a small closed class of so-called *primary verbs*. When functioning as a full verb, BE regularly precedes the subject in *yes-no* questions; HAVE occasionally allows this position in formal BrE; DO always takes DO-insertion:

- (6b) *Is* Jack now fully awake?
- (7b) *Has* the old dancer fond memories of Paris?
- (7c) *Does* the old dancer *have* fond memories of Paris?
- (8b) *Did* her parents *do* nothing to change her mind?

The following central *modal verbs* always function as auxiliaries: *can, could, may, might, shall, should, will, would, must*. Note that they have no base form, only a fixed present and past form. There can never be more than one central modal auxiliary in a predicator. In strings of auxiliaries, the others are typically forms of the primary verbs BE and HAVE.

A predicator is *finite* if it contains a finite verb. A predicator is nonfinite if all the verbs in it are nonfinite. A finite predicator may contain up to three (occasionally four) nonfinite verbs in addition to the finite verb. The distinction between finite and nonfinite hinges on the presence or absence of

present/past marking: a finite verb is either formally present or formally past whereas a nonfinite verb belongs to one of the following three form types:

- (i) infinitives (with or without the infinitive marker): *(to) break, (to) think, (to) worry*, etc.;
- (ii) present participles (*breaking, thinking, worrying*, etc.);
- (iii) past participles (*broken, thought, worried*, etc.).

In the following examples, all the predicators (marked in square brackets) are finite, containing a finite verb (in italics):

- (9) Jack and Jill [*take*] a walk every morning.
- (10) Jack [*takes*] things as they come.
- (11) Jack and Jill [*have taken*] their stand on the issue.
- (12) Both of them [*could have been taking*] the book to the library.

All the non-italicized verbs in these finite predicators (i.e. *taking, taken, have, been*) are nonfinite by themselves. The same is true of *take* when it is an infinitive, not a present form, as in the following examples:

- (13) To *take* a walk would be foolish.

In a string of verbs in a finite predicator, it is the first verb (the first auxiliary) which is finite. This verb is often referred to as the *operator*. To form *yes-no* questions there is *subject-operator* inversion, often referred to as *partial inversion* because only a part of the predicator is moved, cf. examples (2) and (3) above; for discussion see section 5.3 below.

Note finally that there can be only one full verb in a predicator. In examples like the following the second full verb is thus by definition outside the sentence predicator:

- (14) My old friend [*decided*] *to leave* the party.
- (15) His girlfriend [*stopped*] *singing*.

In these examples, *to leave* and *singing* are part of, or fully constitute, the object rather than belong to the predicator (see section 4.3.4).

Let us summarize the defining characteristics of the sentence predicator:

- (i) A sentence predicator is always finite, containing a finite verb, showing formal present/past marking.
- (ii) A predicator contains one, and only one, full verb. In a predicator group, the full verb always assumes head function.
- (iii) Apart from the full verb, a predicator may contain up to three (occasionally four) dependent auxiliary verbs (a modal auxiliary and/or one or more forms of the primary verbs BE, HAVE and DO).

3.2.2. The subject

Once the predicator of a sentence has been found, it is usually fairly simple to locate also the subject. Typically, the subject expresses the person or thing which the predicator says, or predicates, something about. The subject is thus the *topic* of statements, whereas the predicator is part of what is being stated about the subject, the *comment* made about the subject. We can find the subject by asking 'Who or what' immediately followed by the predicator, i.e. 'Who or what P?' The answer to that question is the subject. Consider:

- (1) The parish *vibrated* with gossip the next day.
- (2) It *was* a terrible shock to Mummy and Daddy.
- (3) Daphne *had enjoyed* the illicit character of our relationship.

To find the subject in (1) to (3) we simply ask the question 'Who or what P?':

- (1') Who or what *vibrated*? *The parish* (did)
- (2') Who or what *was*? *It* (was)
- (3') Who or what *had enjoyed*? *Daphne* (had)

While this fairly simple test applies to the vast majority of sentences, there are instances where it does not really make sense to ask 'Who or what P?':

- (4) It *was raining* cats and dogs.
- (4') Who or what was raining? **It*

More formally, the subject displays a number of defining characteristics:

(i) The subject typically precedes the predicator in simple statements (as we see in examples (1) to (4)).

(ii) The subject is always placed between the operator and the rest of the predicator in *yes-no* questions, immediately following the operator (if the predicator is a primary verb it also immediately precedes the subject):

- (1a) *Did* the parish *vibrate* with gossip the next day?
- (2a) *Was* it a terrible shock to Mummy and Daddy?
- (3a) *Had* Daphne *enjoyed* the illicit character of our relationship?
- (4a) *Was* it *raining* cats and dogs?

(iii) Like the predicator, but unlike any other constituent, the subject is always obligatorily present in sentences expressing statements. This means that minimal sentences expressing statements contain S and P only:

- (5) John left.
- (6) The last glimmer of hope evaporated.

(iv) There is *concord* between subject and predicator, i.e. agreement between these constituents in terms of *number* and *person*. With one exception (see below), subject-predicator concord is restricted to the present form of the finite verb: if the subject is in the singular third person (i.e. *he*, *she*, *it*, or anything potentially represented by these pronouns), the verb takes the suffix *-(e)s*, otherwise it appears in its base form:

- (7a) I *take* it easy.
 (7b) She *takes* it easy.
 (8a) The young woman *teaches* English grammar.
 (8b) The young women *teach* English grammar.

The verb BE is especially expressive with respect to concord, being the only verb showing concord in the past form and showing three person distinctions in the present form:

- (9a) I *am* better now than I *was*.
 (9b) You/We/They *are* better now than you/we/they *were*.
 (9c) He/She/It *is* better now than he/she/it *was*.
 (10) The book/books *was/were* far too expensive.

(v) With pronouns to which the distinction between the subjective and objective case applies (e.g. *I/me*, *he/him*, *she/her*, *we/us*, *they/them*), the subjective case is used when the pronoun functions as the subject of a finite predicator (see e.g. (7a-b) and (9a-c)).

(vi) Subjects, but not objects, complements or adverbials can be represented by a pronoun in a so-called *tag question*:

- (11) {Bob} gave them extra work, didn't *he*?
 (12) {You and I} know better, don't *we*?

In some sentences there are *two* subjects, a provisional subject (Sp) and a real subject (Sr). Only *it* and *there* may function as provisional subject:

- (13) It was obvious that he disliked her.
 Sp:pro[It] P:v[was] C:adj[obvious] Sr:cl[that he disliked her]

