

Tense and Aspect
Systems

Östen Dahl

© Östen Dahl 1985
First published 1985
Basil Blackwell Ltd
108 Cowley Road, Oxford OX4 1JF, UK
Basil Blackwell Inc.
432 Park Avenue South, Suite 1505,
New York, NY 10016, USA

All rights reserved. Except for the quotation of short passages for the purposes of criticism and review, no part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher.

Except in the United States of America, this book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, re-sold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

British Library Cataloguing in Publication Data

Dahl, Östen
Tense and aspect systems.
1. Grammar, Comparative and general—Aspect
2. Grammar, Comparative and general—Tense
I. Title
415 P281

ISBN 0-631-14114-6

Library of Congress Cataloging in Publication Data

Dahl, Östen
Tense and aspect systems.
Bibliography: p.
Includes index.
1. Grammar, Comparative and general—Tense. 2. Grammar, Comparative and general—Aspect. 3. Typology (Linguistics) I. Title.
P281.D27 1985 415 85-7531
ISBN 0-631-14114-6

Typeset by Saxon Printing Ltd., Derby.
Printed in Great Britain by The Bath Press, Bath.

Contents

Preface	vi
List of abbreviations	ix
1 General background	1
2 The investigation	36
3 Aspectual categories	69
4 Tense categories	103
5 The Perfect (PFCT) and its relatives	129
6 An overview of the TMA systems of the languages in the sample	154
7 Conclusion	182
Notes	191
References	194
Appendix The TMA questionnaire	198
Index	207

Preface

My interest in tense and aspect goes back at least to 1970. In my early work in this area, emphasis was on the application of notions from logical and philosophical semantics to the analysis of tenses and aspects mainly in English, Russian, and Swedish. At a relatively early stage, however, I felt that a widening of the data base was highly desirable, although it turned out to be quite difficult to find out what the tense-aspect systems of other languages were like from the linguistic literature. This is the background to my switch to a more data-oriented approach.

In 1977, Lars-Gunnar Andersson and I applied for financial support for a large-scale project on 'Universal grammar and language typology', with tense-mood-aspect as one of the fields of research. The Swedish Research Council for the Human and Social Sciences was wise enough to ask us to concentrate on one area. For partly accidental reasons, the subject chosen was the categories of tense, mood, and aspect. Between 1978 and 1982 the Research Council supported the investigation financially. After that time, continued computer processing of the material was made possible by grants from the Faculty of Humanities at the University of Stockholm.

I want to express my gratitude here to those people in Göteborg and Stockholm who were active in the project in one way or another: Karin Aijmer, Lars-Gunnar Andersson, Sally Boyd, Kari Fraurud, Pierre Javanaud, Masha Kopchevskaya, Dora Kós-Dienes, Liisa Karhapää, Kerstin Naüclèr, Maria Toporowska Gronostaj. Dora Kós-Dienes, being the only person ever employed within the project on a more regular basis, should receive special mention: she carried out the bulk of the data collection and analysis work, and I wonder if we would have got through without her unfailing enthusiasm.

Between 1977 and 1985, the role of the computer in linguistic research has changed from being used only by a few specialists to being a necessary tool for almost everyone. In our case, the project was computerized, so to speak, in mid-time. In retrospect, it is hard to see that we could have got anywhere without that. Benny Brodda, Christian Sjögreen, and Carl-

Wilhelm Welin should be gratefully mentioned as those who gave us the necessary help to get the computer analysis going.

Draft versions of the book or parts of it were read and commented on by Bernard Comrie, Casper de Groot, Tore Janson, Dora Kós-Dienes, Nils-Bertil Thelin, and Hannu Tammola. To the extent that I have managed to express myself clearly enough to make myself understood, it is largely thanks to them; none of them should be held responsible for the final result, however.

An investigation which is based on data from as many languages as ours has to rely on the help of many people in different places. In addition to the people otherwise involved in the project, thanks for help in organizing the collection of questionnaires for not so easily obtainable languages are due to Eva Ejerhed, Dick Hudson and Digvijay Singh. In addition, I want to thank the following people, who served as informants and/or investigators for particular languages: Hamid Ahmed Mahmoud (Beja), Fathi Talmoudi (Arabic), Josef Porat and Jan Retsö (Hebrew), Dominic Buttigieg (Maltese), Leo Lindblom and Fessahaie Ghebregzghi (Tigrinya), Tesfaye Alemayehu (Amharic), Valerie Pines, Intizar Salehova (Azerbaijani), Guürel Egecioğlu (Turkish), Chieko Fujio-Düring (Japanese), José Larraín (Quechua), Pedro Monges (Guarani), Barnabas Roberts (deceased) and M.C. Sharpe (Alawa and Bandjalang), John and Ida Wolff (Cebuano), Peter Sengkey (Indonesian, Bugis Makassar), Stephanus Setiabrata (Javanese), No'eau Warner (Hawaiian), Winifred Bauer and Bill Parker (Maori), Shukia Apridonidze, Vilena Jojna, Lily Goksadze, Iza Bakradze, Dali Sakhokia (Georgian), Michael Fortescue and Robert Petersen (West Greenlandic Eskimo), R. Morris and E. Alldrich (Afrikaans), Joyce Hudson and Bernadette Willian (Fitzroy Crossing Kriol), Folke Freund (German), Bo-Lennart Eklund and Michalis Zervov (Modern Greek), Lluís Solanes i Poch (Catalan, Spanish), Jean-Michel Saury (French), Pier Marco Bertinetto (Italian), Gabriella Serban (Romanian), Francisco Lacerda (Portuguese), Anna and Arne Hult and Iskra Jordanova (Bulgarian), Jarmila Panevová and Ivana Seidlová (Czech), Elena Dahl (Russian), Clifford Abbott, Melinda Doxtator, Rebecca Ninham, Mary Jordan and Flora Skenandore (Oneida), Thomas McElwain (Seneca), Elisabet Engdahl and Panit Chotibüt (Thai), Jan-Olof Svantesson (Kammu), Pamela Gichangi (Kikuyu), Maseephu 'Musu (Sotho), Tsokolo Musu (Zulu), Magdalena Wichser, David Sagnon, Oty Sori (Karaboro), A.P. Omamor (Isekiri), Ingela Ökvist and Morakinyo Akintofolarin (Yoruba), Suleyman Njie (Wolof), Liao Qiuzhong, Chen Ping, Wang Juquan and Zhou Huan-chang (Chinese), Diana Krull and Katrin Maandi (Estonian), Orvokki Heinämäki and Marja Leinonen (Finnish), and Istvan Kós (Hungarian).

Over the years, I have had occasion to discuss tense and aspect with a great number of people: it would not be of any use to try to enumerate them all, but thanks are extended to them collectively, as also to all the

people who should have been on the lists above, but who are not, due to faulty memory or book-keeping. Finally, thanks to the staff of Blackwell, for shortening the final phase of this eight-year undertaking by publishing this book quickly and efficiently.

Although the aim of this book is to convey a general picture of what tense-aspect systems of human languages are like, it is not a general introduction to the study of tense and aspect. For this reason, relatively limited attention will be paid to explaining basic concepts and surveying earlier work. Anyone who feels dissatisfied with this is referred to Bernard Comrie's excellent books in the Cambridge Textbooks in Linguistics series (see the bibliography).

Abbreviations

Labels of cross-linguistic category-types

ALREADY	p. 129
CONCL	p. 95
DEFAULT	Default category (p. 19)
DEFAULTd	Default category, dynamic contexts
DEFAULTs	Default category, stative contexts
EXPER	EXPERIENTIAL (pp. 139-44)
FRAMEPAST	pp. 148-9
FUT	FUTURE (pp. 103-11)
FUTi	FUTURE, applies only to IPFV contexts
FUTs	FUTURE, applies only to stative contexts
HAB	HABITUAL (pp. 96-7)
HABG	HABITUAL-GENERIC (pp. 97-100)
HABPAST	HABITUAL-PAST (pp. 100-2)
HABPASTc	HABITUAL-PAST, used also as a counterfactual
HEST	HESTERNAL (p. 126)
HOD	HODIERNAL (p. 125)
IMMED-PAST	IMMEDIATE PAST (p. 127)
IPFV	IMPERFECTIVE (p. 69)
IPFVd	IMPERFECTIVE 'Slavic style' (p. 69)
NARR	NARRATIVE (p. 113)
NOW-TENSE	NOW-TENSE (pp. 95, 176)
PAST	PAST (p. 115)
PASTd	PAST, applied only to dynamic contexts
PASTi	PAST, applied only to imperfective contexts
PASTn	PAST, non-narrative
PASTs	PAST, applied only to stative contexts
PFCT	PERFECT (pp. 129-39)
PFCTq	PERFECT/QUOTATIVE (p. 130)
PFV	PERFECTIVE (pp. 69-89)
PFVd	PERFECTIVE 'Slavic style'
PLPFCT	PLUPERFECT (pp. 144-7)
POSTHOD	POST-HODIERNAL (p. 126)
PRED	PREDICTIVE (p. 110)
PREHOD	PRE-HODIERNAL (p. 125)
PRES-PROG	PRESENT PROGRESSIVE (p. 94)
PROG	PROGRESSIVE (p. 90-5)
PROSP	PROSPECTIVE (p. 111)
QUOT	QUOTATIVE (pp. 149-53)

REM-PAST	REMOTE PAST (p. 175)
RESUL	RESULTATIVE (pp. 133-5)
STAT	STATIVE (p. 29)
VOL	VOLITIONAL (p. 161)

'<' before a category label means that the set of contexts in which the language – specific category is used is included in the normal distribution of the cross-linguistic category, although the fit is not good enough.

'?' after a category label means that the identification is uncertain.

Marking types

M	Morphological
P	Periphrastic
U	Unmarked

Other abbreviations used in the category tables

Adj.	Adjective
Aor.	Aorist
Aux.	Auxiliary
Cont.	Continuous
Cop.	Copula
Dep.	Dependent
Impfct.	Imperfect
Impfv.	Imperfective
Inf.	Infinitive
Pass.	Passive
Pred.	Predicative
Pres.	Present
PrP	Present Participle
PtP	Past Participle
V	Verb

Sentences from the TMA questionnaire are referred to as follows: (Q.nnn) or (Q.nnn:XY), where nnn is the number of the sentence in the questionnaire and XY is the code of the language into which the sentence has been translated.

1

General background

Our ordinary language shows a tiresome bias in its treatment of time. Relations of date are exalted grammatically as relations of position, weight, and color are not. This bias is of itself an inelegance, or breach of theoretical simplicity. Moreover, the form it takes – that of requiring that every verb form show a tense – is peculiarly productive of needless complications, since it demands lip service to time even when time is farthest from our thoughts.

(Quine 1960, 170)

Depending on one's inclinations, one may agree or disagree with Quine concerning the 'tiresomeness' of the property of English he is referring to; it is indisputable, however, that in many languages the speakers are forced by the grammar to pay constant attention to time reference in order to choose correctly among the forms traditionally called 'tenses'. Likewise, it is equally or even more common that speakers have to choose among different forms called 'aspects' and 'moods', the semantics of which tends to be even more elusive than that of tenses.

Tenses, moods, and aspects – henceforth 'TMA categories' – belong to the things in one's native language that one tends to take for granted, and often, they have only attracted the attention of grammarians who have had to explain the use of such categories in one language to speakers of another language in which the system is different. However, since the semantics of TMA categories is connected with concepts that are fundamental to human thinking, such as 'time', 'action', 'event', philosophers have often had occasion to reflect upon their use. In recent years, as the scope of general linguistic theory has widened, there has been an upsurge in studies of TMA systems, where the insights of earlier traditions have been exploited. Still, most such studies build on limited data bases, even when they make universal claims. Some languages have had their TMA systems described in hundreds of monographs and articles; yet it is usually impossible to know to what extent the claims and the

conceptual apparatus of these works can be extended to other languages. Most extant descriptions of the world's languages contain almost no information at all about the use of TMA categories except for the labels that the grammarian has chosen to apply to them. Even if these labels are not just taken over from school grammar – as is often the case – the terminology tends to be too idiosyncratic to warrant proper comparisons with other languages, and the few examples given are more often than not of little help, too.

This book is the outcome of a research project whose aim was to remedy the situation we have just described by creating a data base containing comparable data on the TMA systems of a large number of languages. Thus, the book is a report of a concrete investigation: it contains extensive descriptive material as well as discussions of research methodology. More importantly, however, it aims at conveying a general picture of what a TMA system – primarily, a tense-aspect system – can be like. Expressed in somewhat more ambitious terms, the book is intended to be a contribution to the general theory of tense and aspect, and of grammatical categories in general, based on the analysis I made of data from more than 60 languages collected within the research project. The primary aim of this analysis was to test the hypothesis that the TMA categories that occur in the languages of the world can be reduced to a small set of cross-linguistic category types. In contradistinction to the original project, in which we tried to cover the total TMA field, the final analysis was restricted to tense and aspect categories that occur in affirmative declarative simplex sentences (hence the absence of 'mood' from the title of this book).

A few points about our area of interest should be made right here, although they cannot be developed in detail until later. Even if it is true, as was said above, that TMA categories are linked up with fundamental concepts in human thinking, their study is not co-extensive with the study of temporal, modal and aspectual notions: rather, the former coincides with the latter only in so far as these notions play a role in grammar, i.e. correspond to grammatical categories. If we find that there are cross-linguistic constraints on what notions TMA categories express, plausible explanations to these constraints may be found in limitations on what grammatical categories can be like as well as in properties of human cognitive structures in general. In particular, I shall argue in this book that there are clear differences between those TMA categories that are expressed morphologically and those that are expressed periphrastically, e.g. by auxiliary constructions.

The structure of the book is roughly as follows. In this initial chapter, I shall try to give a general background to what follows, concentrating on some of the fundamental concepts that lie behind the analysis. In chapter 2, I describe in some detail the different stages

of the investigation. In chapters 3–5, the postulated cross-linguistic TMA categories and their manifestations in individual languages are discussed. In chapter 6, a survey of the systems of major tense-aspect categories in the languages of the sample (ordered by genetic groups) is given. Finally, the results of the investigation are summed up in chapter 7.

General semantic and pragmatic considerations

Before going into a discussion of TMA systems, it is necessary to discuss some general theoretical questions which have bearing on our undertaking. Mainly, these are questions of semantics and/or pragmatics: the dividing-line is almost impossible to draw, since the 'meaning' of TMA categories cannot in general be reduced to questions of reference but must be formulated in the broader framework of a theory of language use. When in the following the term 'extension' of a TMA category is used, what is intended is the set of contexts in which the category is found in a language, rather than the set of objects which a term denotes.

Impreciseness and focusing

In this subsection, I shall discuss the two notions of **impreciseness** and **focusing**.

By an imprecise category I mean a category which cannot be defined in such a way that for every member x of its domain (that is, the set of things to which the category can be meaningfully applied), the definition determines a truth-value to the statement that x belongs to the category in question. Instead, even if some members of the domain clearly fall under the category and some clearly do not, there is in an imprecise category a zone in the middle where membership is not clearly defined. In fact, most everyday terms denote imprecise concepts: a classical example is *bald* – it is virtually impossible to define this term in such a way that there will be no unclear borderline cases.

The notion of impreciseness can be extended in such a way as to be applied also to cases where 'membership of a set' is not – at least at first sight – immediately applicable. In particular, when talking about grammatical categories, it may often be more natural to speak of the impreciseness of the borderline between the cases when the category can or should be used (e.g., the cases when, say, the Past tense in English can or should be used) and the cases when it cannot. This will have little or no bearing on the logic of the notion of impreciseness, however.

The notion of an imprecise category or concept has been referred to by several different names in the literature. Philosophers usually refer to the notion by the term 'vagueness', although this probably does not correspond to the most common use of the word *vague* in everyday language (to say that someone expresses himself in a vague way usually means that he does not give very specific information on a subject). Recently, the term 'fuzziness' has become popular in connection with the development of 'fuzzy set theory'. Below, I argue that 'fuzzy set theory' does not provide us with an adequate logic for imprecise concepts and I outline an alternative treatment.

Another notion that has played an important role in recent discussions of cognitive structures is that of a 'prototype'. Underlying this notion is the idea that concepts are best understood in terms of a description of what the 'best exemplar' of the concept or category is like. There is a clear and direct relation between this and imprecise categories: speaking of 'the best exemplar' presupposes that not all members of the category have the same status – the extension of a category has to have a 'focus' and a 'periphery', where those entities that belong to the periphery will have a more or less dubious membership. In fact, one could say that the weakest formulation of the main claim of prototype semantics is just that concepts are generally imprecise. A stronger version of the theory would say that there are differences in status even among those entities that are clearly within the extension of the category. For instance, it has been argued that members of a category like 'birds' differ in 'typicality' – sparrows are more typical birds than penguins, although there is no doubt that a penguin is a bird. We shall therefore introduce the term **focused** category for those categories the extensions of which have a definable focus (or alternatively, several foci), admitting, however, that this property will in most cases coincide with impreciseness.

We shall now discuss in more detail, and with some degree of formality, the underlying logic of impreciseness and focusing.

Suppose we have a universe of discourse U consisting of a set of persons, and consider some predicates that we could use to describe the members of this set. To start with, consider a description such as ' x is male'. For most practical purposes, this can be regarded as expressing a precise concept: that is, for every member of the set, it is possible to state definitely whether he/she is male or not. Let us now do the following: (i) assign one of the numbers 1 and 0 to the members of the universe of discourse, accordingly as the predicate under discussion is true of them or not; (ii) order the members of U in such a way that those who have been assigned 1 precede those who have been assigned 0. The result might be depicted as in figure 1.1.

If we have an imprecise predicate – say *bald* – on the other hand, there will be a subset of the universe of discourse which cannot be assigned 1 or 0, at least not for the time being. Thus, we cannot use

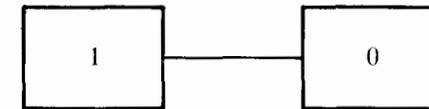


Figure 1.1

figure 1.1 as a representation. We might instead use figure 1.2, where the unclear cases are in the middle.

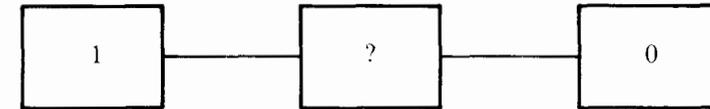


Figure 1.2

Assume, for the sake of discussion, that the only criterion for judging the baldness of a person is the number of hairs on his head, but that there are different options as to where the borderline goes, that is, that the maximal number of hairs that a bald person can have is not defined (although it is known to be greater than zero). (These are the assumptions that underlie the old 'Paradox of the Bald Man'.) In such a case, we know that if we decide that a person x is bald, we must also assume, in order to be consistent, that any person who has more hairs than x is also bald. What this means is that we can break up the middle group into smaller ones, which are ordered relative to each other in such a way that assigning baldness to a group entails assigning baldness to all other groups to the left of it (see figure 1.3).

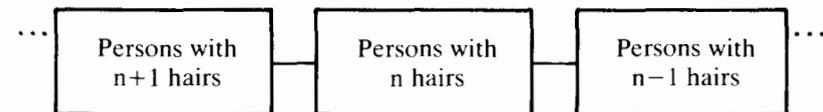


Figure 1.3

One important point about the kind of impreciseness that we are discussing is that we can reduce it by choosing a more precise concept, i.e. we can assign a truth-value to some or all of the elements of the set C . Let us say that a concept Q_1 is a **sharpening** of concept Q_2 if all elements with determinate truth-values relative to Q_2 have the same truth-values relative to Q_1 and in addition at least one element with undefined truth-value in Q_2 has a determinate truth-value in Q_1 . Two possible sharpenings of the concept of baldness are shown in figure 1.4.

We see that the diagram so to speak shrinks when the concept is sharpened. We shall give a more formal characterization of what happens shortly. First, however, let us introduce the distinction

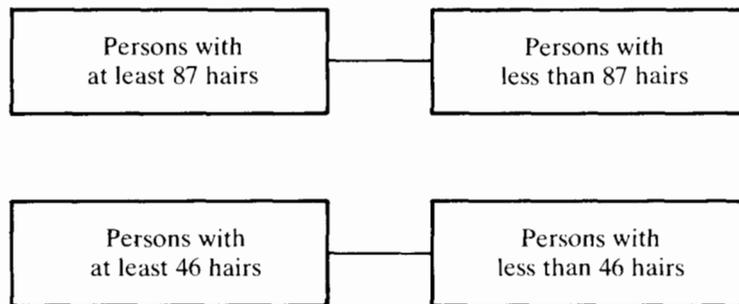


Figure 1.4

between **one-dimensional** and **multi-dimensional** impreciseness. Baldness, understood as above, is a paradigmatic case of one-dimensional imprecision – the imprecision consists in an indeterminacy as to where a borderline is to be put on a one-dimensional scale. Consider now instead the concept of ‘being Swedish’. There are of course plenty of clear cases: on one hand, persons who are born and live in Sweden, who speak Swedish and have Swedish parents etc.; on the other, people who have nothing to do with Sweden at all. But consider e.g. a person who was born in Sweden of Swedish parents but who has lived in the United States all his life, is a US citizen and does not remember a word of Swedish. We immediately see that the problem here is of another character: we do not know which of a number of criteria should be decisive, and there is no way of reducing ‘Swedishness’ to a one-dimensional scale. Nationality words are thus paradigmatic cases of multi-dimensional impreciseness.

Discussions of impreciseness, whether under the name of vagueness, fuzziness or whatever, often concentrate on one-dimensional impreciseness, although multi-dimensional impreciseness is at least as interesting, and furthermore can be regarded as the more general concept, of which one-dimensional impreciseness is just a special case.

In the case of multi-dimensional impreciseness, we need a more complex representation than what we had in the simple case exemplified in figure 1.4. In order not to complicate things more than necessary, let us confine ourselves to two-dimensional cases. If we simplify the concept of being Swedish in a way similar to what we did with baldness, we might assume that the only two criteria that are of importance for judging a person’s nationality are his present citizenship and his place of birth. Let us call the propositions that each of these criteria hold p and q , respectively. We then get four logical possibilities, which we can denote by p_1q_1 , p_1q_0 , p_0q_1 , and p_0q_0 , according to the respective truth value of p and q . Of these, the cases where both criteria go the same way – p_1q_1 and p_0q_0 – are clear cases:

in the first, the imprecise category holds, in the second, it does not. The other two cases – p_1q_0 and p_0q_1 – constitute the fuzzy area: if someone is a Swedish citizen but was not born in Sweden – or vice versa – we could decide either way. Since each of the criteria can be accepted or rejected independently of the other, they cannot be said to be ordered relative to each other, and the diagram would have to look like figure 1.5.

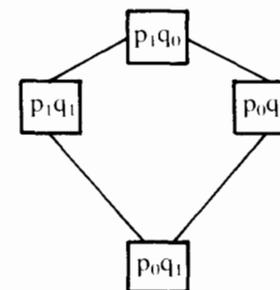


Figure 1.5

This structure is a simple example of a formal object called a **lattice**. (In fact, figures 1.1–2 are also lattices, but of too trivial a kind to be good illustrations of the concept.) Lattices are well-known structures in mathematical logic. They – or more specifically, the kind of lattices known as **Boolean algebras** – can be used to represent e.g. both set-theoretical and truth-functional relations. I shall claim without formal proof here that the logic of imprecise concepts can also be treated in terms of Boolean algebras, with the nice consequence that all the familiar properties of such algebras can be assumed in the discussion.

Let us now return to the concept of sharpening. Call the category in figure 1.5 C_0 . Figure 1.6 then represents the possible ways of successively sharpening C_0 .

We are now in a position to fulfil the promise to give a formal characterization of a sharpening. In terms of lattice theory, a sharpening of a concept A into another concept B is a **homomorphism** from the Boolean algebra representing A into the Boolean algebra representing B . A homomorphism, basically, is a function which preserves the relations between the elements in the algebra. As we have already said, sharpening a concept means that the lattice that represents it ‘shrinks’. The final result will always be the minimal two-element lattice of figure 1.1 which represents a precise concept.

Looking closer at the sharpenings in figure 1.6, we see that all the sharpenings have as a consequence that C_0 is reduced to a one-dimensionally imprecise category. Of particular interest are the lattices C_1 –4, in which the cells 2 and 3 are ordered with respect to

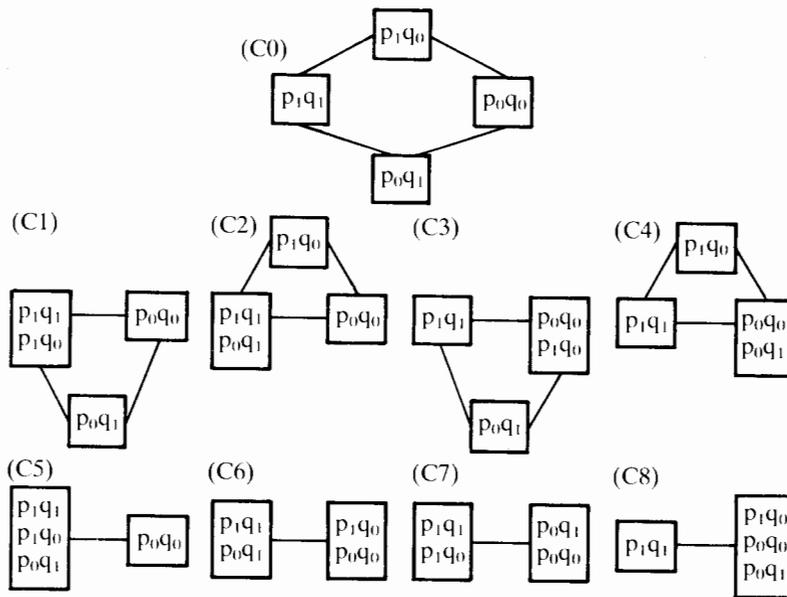


Figure 1.6

each other. We shall say that in such situations, one criterion is **dominant** relative to the other. More precisely:

a property p is dominant relative to a property q in the category C iff no entity can be included in the extension of C which has q but not p .

A concrete example will illustrate what this will mean in practice. Suppose for example that C_0 is the interpretation of *Swedish* and that p and q mean 'is a Swedish citizen' and 'was born in Sweden', respectively. As long as we have not sharpened C_0 in any way, these criteria are equally important. If we go from C_0 to C_4 , however, the result is that the criterion 'is a Swedish citizen' becomes dominant in the sense that it is not possible to include in the extension of C_0 any individual that does not have this property: we cannot for instance sharpen C_4 into C_5 . Another way of expressing this is to say that 'being a Swedish citizen' becomes a necessary condition of being Swedish. It may still not be sufficient, though: we may sharpen C_4 further into C_8 , which means that both p and q are necessary for C_0 : in order to be Swedish, you have to be both a Swedish citizen and born in Sweden. In the same way, we may sharpen C_0 into C_3 , making 'was born in Sweden' the dominant criterion.

Proceeding now to the problem of describing focused categories, we may note that it is often assumed that we can characterize the focus of a category in terms of a list of prototypical properties. For the concept 'bird', for instance, such a list might include 'having wings',

'having feathers', 'being able to fly' etc. The discussion above suggests that the members of such a list of categories may not have equal status – some may be dominant relative to the others. As a concrete example, take the word *man* in English. In its prototypical use, it denotes male adult human beings. We might then suggest that the list of properties corresponding to the prototype of *man* consists of three elements: 'male', 'adult' and 'human'. The extended uses of *man* are characterized by subsets of this list of properties: for instance, *man* may mean 'male human', as in *Man is usually physically stronger than woman*, or just 'human', as in *All men must die*. But it does not seem natural to use *man* in the sense 'grown-up human'. Thus, the properties 'male' and 'human' are dominant relative to 'adult' in the meaning of *man*. We shall see that the concept of a dominant parameter is often relevant in the description of TMA categories. For example, the category Perfective will usually be interpreted as 'perfective' and 'past', although with the first feature clearly dominant.

Let us now briefly return to the theory of fuzzy sets. Fuzzy set theory is a version of set theory which takes membership in a set to be a matter of degree rather than a binary, yes/no question. In technical terms, the characteristic function of a fuzzy set may take any value on the scale between 1 and 0, rather than just the endpoints of that scale. Applying this to imprecise terms, one might suggest that the borderline cases are those where the membership in a category has an 'in-between' value. Thus, someone who is Swedish by birth but not by citizenship might be assumed to belong to the category of Swedes to a degree of, say, 0.75. It seems to me that the idea of explaining imprecision by a theory of this kind is essentially misguided. Its basic flaw is that it confuses indeterminacy with graduality. If one assigns an absolute value to e.g. the Swedishness of the person in question, one creates precision where there is none, and thus entirely misses the point.

The notion of 'basic meaning'

The notion 'basic' (alternatively 'primary') 'meaning' (alternatively 'sense' or 'use') can in fact be interpreted in several ways. To start with, we can look at it either extensionally or intensionally. In the first case, we divide the extension of a term into different regions, one of which we – for whatever reason – look upon as 'basic' or 'primary' with regard to the others. A case in point is the postulation of 'focal' or 'prototypical' uses which are contrasted against 'peripheral' ones. In the intensional case, we might for instance postulate that the 'meaning' or 'sense' of the word consists of several components (features, markers or whatever), one or more of which are then said to be primary or basic with regard to the others. This is the approach we

have taken when we have talked about 'dominant' parameters. It is of some importance to keep these two approaches apart, since – particularly in discussions of TMA categories – the choice of one before the other may very well lead to different consequences. For instance, we shall argue that although the primary – in the sense of prototypical – use of the Future in English involves both 'future time reference' and 'intention', the basic meaning in the sense of dominant parameter, however, is 'future time reference'.

Another difficulty with the notion 'basic meaning' is illustrated by the following example. The extension of *cat* would normally be said to be something like the set of all cats. However, when *cat* is used generically, it might be regarded as referring to a kind or species. Given that the borderline for what is a cat is imprecise, we might get several possible 'kinds' that *cat* refers to, and *cat* might then be taken to be ambiguous. Thus, the kind might be taken to be the species *Felis felis domesticus*, but it could also be the family that comprises lions and tigers (as in *lions and other big cats*). It would appear natural to say that the former possibility constitutes the 'basic' or 'primary' alternative – it also coincides with the narrowest sharpening which contains only the most prototypical exemplars. However, it is not always self-evident that the 'primary' meaning can be identified with the narrowest delineation of a concept. Consider the word *dog* in English, which may be taken to refer either to the species *Canis canis* or to the male members of that species. If the situation were quite parallel with that of *cat*, one would classify the narrowest reading – 'male dog' – as the primary meaning of the word *dog*. However, this appears rather counterintuitive: most people would probably feel that this interpretation is secondary relative to the gender-neutral one. The point of the argument is that we may have to identify a 'basic sharpening level' for a concept, which may not be the narrowest possible one, as a simple identification of 'primary meaning' and 'prototypical cases' would imply.

Secondary meanings

Given the notion of a focused category, a 'secondary meaning' ('sense, use, reading, interpretation') could be defined negatively as something that is outside of the focus. (Obviously, we are now taking an 'extensional' perspective in the sense defined above.) If we take the focus to be defined by a set of properties, then the prototype, secondary meanings would, in the simplest cases, be represented by a subset of the prototype, that is, some of the prototypical properties would be present and others not. Secondary meanings would thus be weakened primary meanings, so to speak. Thus, when we talk about lions as being cats, we might be said to see lions as objects that are cats

in a secondary sense since they have only some of the properties of the focal cats. This may be a possible way of accounting for some kinds of ambiguities: I do not think that we can adequately describe typical cases of polysemy this way, however. Rather, the formation of polysemic items should be seen as an active process, which eventually leads to the creation of what could be called **secondary foci** which are characterized by sets of properties containing elements not present in the original prototype. For instance, a film star is no doubt so called because there is some similarity with a star in the primary sense: but the intersection between the properties of a film star and a star in the sky is hardly sufficient to account for how we understand the word 'star' in a film context. The traditional concepts of metaphor and metonym will cover a significant part of this, but there are no doubt other processes.

One powerful mechanism for creating secondary foci and secondary interpretations is what we can refer to as the **conventionalization of implicatures**. Following what is by now standard terminology, I use the term **implicature** (coined by the philosopher H. P. Grice – see Grice 1975) to mean something that can be inferred from the use of a certain linguistic category or type of expression, although it cannot be regarded as belonging to its proper meaning. It should be noted that given a prototype approach to meaning, the borderline between implicatures and meaning proper is much less clear than it may be in other theories, since a prototype is a set of 'characteristic' rather than a set of 'defining' features. What happens when a conversational implicature is conventionalized may be described as follows: if some condition happens to be fulfilled frequently when a certain category is used, a stronger association may develop between the condition and the category in such a way that the condition comes to be understood as an integral part of the meaning of the category. For instance, the tendency for categories like the English Perfect to develop 'inferential' interpretations might be explained in this way (see chapter 5 for further discussion). Another example would be the development of Perfects and Pluperfects into recent and remote pasts, respectively (see chapter 5).

Accidence categories and Gricean principles

In the preceding sections, I have discussed some general problems of semantic description. I now want to turn to issues specific to a restricted class of categories that are exemplified in many if not most natural languages. This class includes categories such as number, gender¹, 'level of formality', but most importantly for us, TMA categories such as the Past tense in English. I shall not try to give precise criteria for what should be included in this list, but shall argue

that the categories I have enumerated have certain interesting common properties that warrant treating them under one heading. There is no standard term for them in contemporary linguistics: 'grammatical category' would be a possibility, but would most naturally be interpreted much more widely. Characteristically, the categories I have mentioned are those that are most often expressed morphologically; it is therefore tempting to refer to them as the class of 'inflectional' or 'morphological' categories; this would be misleading, however, since they may also be expressed in other ways, e.g. by syntactic means, and there may also be inflectional categories which we would not like to include here. In traditional grammar, the term 'accidence' was used for categories that characterized expressions (words, mainly) 'accidentally' or 'contingently' as opposed to 'inherent' or 'essential' features of lexical items. Since this term has lately gone out of fashion, I feel free to usurp it for my purposes; consequently, I dub the categories enumerated above 'accidence categories', preferring this expression to the potentially misleading 'accidental categories' and the clumsier 'categories of accidence'.

Typically, accidence categories can be said to work in the following way. There exist in the language alternative forms, the choice between which is regulated by some parameter pertaining to properties of the objects or situations referred to in the utterance, or of elements of the speech situation, or the relations between the former and the latter. Sometimes, the choice may be dependent upon some feature of the linguistic context. The choice between the alternative ways of expression is typically a 'forced one'; not only in the sense that you have to choose one of the alternatives (or to remain silent) – something that is of course always trivially true when you have to choose between two ways of saying something – but also in the stronger sense that in choosing a certain form you voluntarily or involuntarily convey a piece of information, viz. the information that the conditions for the appropriate use of that form are fulfilled, and at the same time you make a commitment – however implicit – to the truth of that information.

As a simple example, consider the use of the category of natural gender in a language like French. As is well known, every adjective in French must agree with its head noun (if it is attributive) or controlling noun phrase (in other positions). This means that whenever a speaker uses an adjective about a person, he has to make a decision about the sex of that person. Thus, a sentence such as (1.1) carries the information that the speaker is female.

(1.1) Je suis contente 'I am satisfied'

When speaking, we convey information of various kinds, and in various ways. Consider the difference between saying (in English) I

am from Sweden and making an arbitrary utterance, say *It's raining*, with a Swedish accent. In both cases, the speaker can be said to convey the same piece of information: that he is from Sweden. In the first case, it is part of his 'intended message': he has himself chosen to make the statement in question with the intention that the addressee should believe that it is true. In the second, on the other hand, it is information that he conveys whether he likes it or not, in view of his imperfect knowledge of English. (We disregard for the time being the possibility that the Swedish accent is intentionally faked or exaggerated.) These are clear examples of the distinction between intentionally and unintentionally conveyed communication. (For a more careful taxonomy of the field, see Allwood 1976.)

Returning to (1.1), we see that the compulsory character of the use of feminine gender for women means that you cannot avoid conveying this kind of information, whether or not it belongs to the 'intended message', i.e. the information that you want to convey. This is so in spite of the fact that, of course, the choice between different grammatical forms is something that can be more easily manipulated than the phonetic features of speech referred to as 'accent'.

The example I chose to illustrate the compulsory character of accidence categories did not concern a TMA category, but what I have said applies with equal force to them. The 'semantic irrelevance' of TMA categories is particularly striking in a typological survey like the present one; in spite of the great similarities between TMA systems in different languages, and the obligatoriness of language-specific categories, there is hardly any distinction in the TMA field which is marked in all languages. Still, in most cases the loss of information entailed by not marking a potential distinction seems to have little or no negative effect on communication. Of course, there are usually optional means of making a distinction, if needed: in practice, however, these means need only be used in fairly infrequent cases.

Formal interpretation rules for natural languages are most commonly given in terms of truth-conditions. Such rules are of the general form:

(1.2) A sentence S is true iff p.

It should be clear from what we have said that the semantics of accidence categories can only partly be accounted for in terms of such truth-conditions; the choice between different ways of marking a sentence by such categories may be only indirectly related to the question of whether it is true or false. For instance, the choice of the Present Perfect instead of the Simple Past in English may render the utterance unacceptable but rarely makes it false. Even if it would be too strong a statement to say that TMA categories have nothing to do with truth-conditions, I prefer to think of them as governed in general

by 'conditions of use'. I also prefer to speak of the 'use' rather than the 'meaning' of TMA categories, since it appears to me that the latter should be restricted to what directly concerns 'the intended message'.

Linguistic communication is commonly supposed to obey some principles like the Gricean conversational maxims (Grice 1975), in particular, his 'maxim of quantity' and his 'maxim of relation':

Maxim of quantity: Make your contribution as informative as is required (for the current purposes of exchange).

Maxim of relation: Be relevant.

We can see that in a sense (1.1) violates both these principles: it contains information about the sex of the speaker, although this is normally redundant in the speech situation. I think it would be wrong to say that the existence of such cases invalidates the principles, but it is essential to keep in mind that it is a pervasive trait of human languages that they force us to make our messages 'too informative' in various respects, and that accident categories play an extremely important role in this regard.

Having made this claim, I hasten to make an important modification. I have used the word 'obligatory' several times. It turns out, when you look closer at the facts, that this word is often too strong. For instance, it is tempting to say that the Past tense in English is obligatory in the sense that whenever you talk about the past, you have to use it. However, as is well known, there are certain styles or certain situations when it is perfectly possible to use the Present tense about the past:

(1.3) Suddenly this guy comes up to me and says . . .

We would thus have to modify our statement to say that we normally mark past time reference by the Past tense or that such marking is the 'default' case. This suggests that the characteristic property of accident categories may be that they are used systematically rather than obligatorily. I think that the fact that most languages seem to have systematic marking of certain semantic features is a non-trivial fact in need of an explanation. In particular, I think it is a challenge to developmental psycholinguistics to explain how it comes about that such principles of systematic marking are usually willingly accepted and learnt (at least seemingly) without great trouble by children at the age of two or three.

One might speculate that we in fact have a need to disregard the Gricean maxims of relation and quantity, resorting instead to more automatized principles of what information to include in the message. In fact, a person who really wants to follow the principle of never saying anything redundant will have to devote a lot of energy to

checking what he is saying. Even if we do not go to such extremes, it is clear that there are quite a few different factors that come into play in determining what is optimal in conversation.

Roman Jakobson (and perhaps someone before him) said somewhere that languages do not differ so much in what they can express as in what they must express. If this is true of anything, it is true of accident categories. A direct consequence is that we can draw no conclusions from the non-existence of a certain accident category in a language or an idiolect about the cognitive capabilities of its speaker(s).

One consequence of the above-mentioned properties of accident categories should be mentioned. The fact that the semantic features involved in accident categories typically do not belong to the 'intended message' makes it rather difficult to arrive at clear judgements of how many 'readings' one should assign to them. Even if one can isolate a set of factors that influence the choice between two forms, it is not obvious that one is thereby entitled to regard the forms in question as ambiguous. For instance, the choice between the second person pronoun *tu* and *vous* in French depends on at least two factors: (i) the number of persons one is addressing, (ii) the degree of formality of the relation between the speaker and the addressee(s). The question then is: is *vous* ambiguous between 'plural' and 'formal' or perhaps even three-ways ambiguous: 'formal singular', 'informal plural' and 'formal plural'? If we accept the not too implausible idea that *vous* is ambiguous, what then about the English *you*, which can be used in all the cases mentioned but also when it corresponds to French *tu*? Clearly, to resolve such questions, we would have to take into account rather abstract considerations such as the general relevance of the category in question in the language. Thus, a possible position would be that *you* is ambiguous between singular and plural but not between 'informal' and 'formal' since the latter category is not systematically marked in English.

It may be noted that the assumption that there is a universal level of semantic representation at least in its stronger versions, where it is coupled with the assumption that translatability from one language into another implies identity of semantic representation, leads to the consequence that whenever an expression has more than one translation into another language it is ambiguous. Whereas such an assumption can be shown to have rather absurd consequences – any English sentence would e.g. be ambiguous between three or more levels of formality, in view of multiple translations into languages like Japanese where such a number of levels are systematically distinguished – similar claims have been made for e.g. the semantics of tense and aspect, where they are much more difficult to refute.²

Different ways of accounting for grammatical categories

Suppose that we have a language where nouns can have two grammatical numbers, Singular and Plural. Let us consider some different possibilities as to the status of the category of number in the grammar of such a language.

- (i) We might simply say that there is a parameter Number which has the two values Singular and Plural. Every noun in a text will then be characterized as being either Singular or Plural.
- (ii) We might postulate a binary feature, say Plural, which has two values, '+' and '-'. Again, every noun form would be either +Plural or -Plural.
- (iii) We might regard Plural as a 'flag' or 'marker' characterizing certain nouns. Singular nouns would then simply be those which lack this flag.
- (iv) Plural could be regarded as an 'operator' which applied to a (singular) noun creates a plural one.

It may not be quite easy to see how these alternatives differ from each other in practice – they may appear to be more or less notational variants. However, which one we choose will in fact be crucial for the claims we want to make about the functioning of the grammatical system of the language. First, we may note that (iii) and (iv) differ in one clear respect from (i) – with (ii) as a somewhat equivocal possibility in between – by ascribing a clear asymmetry to the two members of the opposition, since in the former, we treat Singular as being more basic – as the 'unmarked' or 'default' member of the opposition. In the terminology of European structuralism, (i) treats Number as an 'equipollent' opposition, whereas the others treat it as a 'privative' one. Another difference would be that (i) is easier to reconcile with the possibility of having more than two values.

In set-theoretical terms, any of the above alternatives would be equivalent to the postulation of a function from noun forms to something, but this 'something' would be different. The simplest case is (iii): it would involve a function from nouns to truth-values, or alternatively, to the integers '1' and '0', which in its turn is equivalent to a one-place predicate. Case (ii) could be interpreted in the same way, or as equivalent to (i): as a function from nouns to any delimited set of objects.

As we have seen, (ii) and (iii) could be regarded as equivalent. One thing that seems to distinguish them, however, is the possibility of interpreting (ii) in such a way that one allows for cases where the value of the feature is 'zero' or 'not defined'. In set-theoretical terms, this corresponds to the distinction between a total and a partial function: the question is whether the function has a determined value for all members of its domain or not. In fact, (iii) may also be interpreted so

as to allow for partial functions, if we put a restriction on the set of objects to which the flag can be applied.

(iv) is different from all the others in that its co-domain is, like its domain, a set of expressions: it must be regarded as a function from word forms to word forms. The importance of this is that the logic of (iv) is rather different, and in crucial ways more powerful, than that of the other alternatives. Suppose that the grammar defines n binary features or flags for a certain type of expressions. There will in such a situation never be more than 2^n possible combinations of those features, in the same way as a set of n one-place predicates will never yield more than 2^n possible descriptions. A set of operators, on the other hand, can well give rise to an infinite set of objects, since they can potentially be applied recursively, that is, it is in principle possible that an operator gets its own output as input. Also, two operators may apply to one and the same thing in different orders, yielding different outputs.

Thus, whereas the alternatives (i–iii) are representable in terms of monadic predicate calculus, i.e. to a version of predicate calculus with only one-place predicates, alternative (iv) demands a more powerful logic, which we shall refer to as 'operator' logic.

The possibility of representing a theory in monadic predicate calculus is in fact of great importance from the point of view of axiomatization, since monadic predicate calculus, as distinct from predicate calculus in general, is decidable. In principle, this means that we can regard the system as having the properties of an even simpler logic, viz. propositional calculus. Propositional calculus can be regarded as a kind of Boolean algebra, and we can therefore refer to categories that are thus reducible as 'Boolean'. The Boolean character of TMA categories in a binary feature framework is clear already from the fact that since such categories normally pertain to whole sentences, they would, if treated as predicates, be 'zero-place' rather than anything else, and thus be equivalent to propositional variables.

We may illustrate the difference between the 'one-place predicate' and the 'operator' alternatives by making a short digression on the semantics of adjectives. Early treatments of the formal semantics of natural languages translated adjectives into one-place predicates *tout court*. Combinations of adjectives, such as in *small red house*, would then have to be treated as equivalent to conjoined constructions, e.g. *small and red house*, which implies that the order of the adjectives is not important. However, for many cases, such a treatment is not adequate: for instance, *Italian fake money* might be e.g. counterfeit money manufactured in Italy, regardless of what country it is supposed to be from, whereas *fake Italian money* is counterfeit liras, wherever it is made. To account for such cases, later works (see e.g. Kamp 1975) have treated adjectives as operators, i.e. as functions

from common noun phrases to common noun phrases.

Binary feature models have been extremely popular in recent linguistics, inter alia for the description of TMA categories (cf. e.g. Haltof 1968, Pettersson 1972, Thelin 1978). It is therefore important to consider the inherent limitations of these models³ and the restrictions they put on what you can do with your system.

If we look at tense logic, which is the kind of logic commonly supposed to mirror the behaviour of tense categories, it is easily seen that it is not reducible to a binary feature model. Tense logicians employ operators such as 'it has been the case at some point in the past that', 'it will be the case at some point in the future that' etc. Complex formulae involving the nesting of several such operators are readily constructed. The translations of these formulae into natural languages tend to be rather unnatural, but it is at least not too hard to find pairs of natural language expressions which differ only in the order in which TMA categories have been applied, e.g.:

(1.4) Many people *have been going to marry* Susan.

(1.5) John *is going to have married* her next week

It thus seems probable that a binary feature model, or anything logically equivalent, cannot account for TMA systems in general. Still, it is of course not excluded that parts of these systems may be described within such a restricted system. In particular, we may note that there seems to be a strong correlation between the possibility of having 'nested' structures and the ways in which the categories in question are marked. Categories of the type traditionally labelled 'inflectional' seldom behave like operators: usually, an inflectional category is applied to a word only once, and order of application does not matter: it is rather improbable that the plural of an accusative noun would be different from the accusative of a plural noun. As can be seen from (1.4-5), periphrastic constructions are not limited in the same way: at least in principle there are no constraints on their nesting possibilities. To a somewhat lesser extent, the same can be said of derivational categories: there is e.g. nothing that disallows nominalizing a denominal verb – in fact, the word *nomin-al-iz-ation* is a case in point. Now, TMA categories are expressed by both inflectional, derivational and periphrastic means, and it can thus be expected that there will be differences in how easily they nest.

But there may also be differences in their need to nest: some categories may have a semantics that is 'binary feature-like' and others may not. If we suppose that languages tend to optimize the expressive power of its categories, we would then predict that categories with a non-Boolean semantics are in fact more often expressed by periphrastic and derivational means than other categories. 'Every category has the marking type it deserves.' We shall have occasion to test this possibility in the course of the investigation.

It is important to note that what we have been talking about in this section are the limitations of pure binary feature models: that is, models that do not involve any structures that go beyond simple unordered sets of binary features. Any additions to that, such as the introduction of orderings of features or hierarchical relations of any kind, may well enhance the logical power of the system so as to make it 'non-Boolean'. Since at least some of the works mentioned above involve such extra structure, it is not clear that what we say here can be applied to them, and this section should not be seen as providing direct criticisms against these works but rather as an attempt at clarification.

Markedness

The concept of 'markedness' has by now a rather long and complicated history in linguistics, which I shall not try to go into. In the paradigmatic cases, a grammatical opposition consists of a zero-marked member with less specific interpretation which is opposed to an overtly marked member with more specific interpretation, where the overt marking involves the addition of an extra morpheme. English genitive formation could be an example of this:

John: John's

From such clear cases the concept of markedness has been extended in various ways, a development which has gradually led to a situation where the connection with the original, concrete use has become rather weak and where it is assumed that almost any linguistic choice has a 'marked' and an 'unmarked' alternative. I feel that there may be a certain point in terminological conservatism here, and would therefore like to make a distinction between 'unmarked category' and 'default category'. An unmarked category would be such a member of a grammatical opposition that has the less complex or (relative to a given grammatical description) the basic or non-derived form. The term 'default' – well-known from computer terminology – will be used instead of 'unmarked' to express the more general idea of being the alternative that 'is felt to be more usual, more normal, less specific than the other' (Comrie 1976, 111). In other words, 'marked' would be understood as concerning the expression or form of a category, whereas 'default' would rather pertain to its meaning or use.

This usage has among other things the advantage that we can take care of 'paradoxical' cases of marking, such as the person endings in the Present tense of English verbs, where the 3rd person is marked in our sense, since it has the ending -s, although it can reasonably be argued to be the default choice relative to the 1st and 2nd persons.

Some important notions in the study of TMA systems

Definitions and impreciseness in science

Impreciseness, which was discussed at length above, is a property of almost all notions or categories used in everyday thinking, and terminological discussions in science often aim at reducing this impreciseness. This is a laudable and necessary activity, but may not be possible to carry to its logical limits. Impreciseness is not always an entirely bad thing: it may be that in order to get rid of it we have to make arbitrary decisions that in fact make the notions we use less useful. Suppose we have two phenomena, A and B, each of which is typically characterized by a set of properties. There may then be various borderline cases between A and B which share some of the properties of A and some of the properties of B. Ideally, we would like to make our notions so precise as to be able to assign each of these borderline cases to either A or B. However, by doing so we may miss the fact that the borderline cases are interesting precisely because they are like A in some respects and B in other respects. Also, they may differ between themselves in having picked out different subsets of A's and B's properties: forcing some of them into A and others into B obscures the fact that they are all somewhere in between. For instance, everyone can tell the difference between a man and a chimpanzee: that we are not able to make up our mind whether some fossil should be regarded as a human or an ape does not diminish the value of the distinction. Similarly, refusing to answer the question whether the English Perfect is a tense or an aspect does not mean that one does not know what tense is and what aspect is.

I think that most concepts are learnt by some kind of ostension: a typical exemplar of the concept or category is pointed out and the learner abstracts from it a prototype, i.e. a list of typical and salient properties of the category. I believe this to be true also of many concepts used in scientific thinking: we learn what a 'subject' is by being exposed to simple examples such as 'Socrates' in 'Socrates runs'. Similarly, we know – or think we know – what a 'tense' is because we know what the Past tense in English is like. Now, in spite of its usefulness, ostension obviously has its pitfalls: if the learner is unlucky, the purportedly typical exemplar turns out to be some kind of very infrequent and freakish variety of the category – the child who meets a Great Dane as his first canine creature may have serious trouble with his concept of dog later on. In the same way, a linguist who studies one language or a couple of languages from a restricted area may be unlucky enough to meet grammatical phenomena that turn out to be very untypical from a universal point of view. An investigation of the kind presented in this book may, if it is successful, sort out the Great Danes from the more common mongrels in our

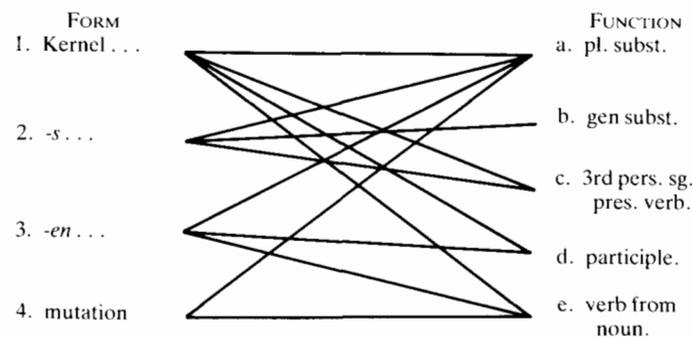
linguistic repertoire. I shall later argue that at least one assumedly 'paradigmatic example' of a TMA category, viz. the perfectivity-imperfectivity distinction in Slavic, is a rather peculiar animal in various respects.

The notions of 'category' and 'grammatical category'

The most fundamental terms are usually the most difficult ones to define. The word 'category' has a wide range of uses in linguistics and related branches of science: this makes it hard to employ it without giving readers various relevant and irrelevant associations. However, there are few alternatives to the term as a label for the units that build up TMA systems, that is, things like the Simple Past in English or the Imparfait in French. Already at this stage some confusion may arise: categories may be defined on different levels, and we could equally well choose to talk of e.g. 'the category of tense' or 'the category of aspect'. Since it is important to our undertaking to be able to treat the individual elements of a TMA system as entities in their own right, the former alternative will be preferred here.

As we have already stressed, the categories that we are primarily interested in are those that play a role in grammar – what could naturally be called 'grammatical categories'. This term, again, is not unproblematic – first, it is also very wide, and could equally well include things like 'subject' and 'object' or say, 'reflexive pronoun', as the categories that interest us here. One might think that 'morphological category' would be more adequate – the Simple Past in English clearly is something that belongs to morphology, but as we shall argue in more detail later, this would delimit our area of study too much. The term 'grammatical category' is problematic from another point of view, too, however: language is commonly thought of as consisting of a level of expression or form and a level of content or meaning, and it is tempting to think that any entity that figures in a linguistic theory must belong to one of these. The view I shall take here, however, is basically that expressed already in Jespersen 1924, viz. that we must assume that grammatical (or as he says, 'syntactic') categories are 'Janus-like' in facing 'both ways, towards form, and towards notion': they 'form the connecting link between the world of sounds and the world of ideas'. The example Jespersen uses happens to be one mentioned in the preceding paragraph, viz. the Simple Past (Preterit) of English, as illustrated in figure 1.7 (Jespersen 1924, 56).

I part with Jespersen, however, in taking the view that it is possible to arrive at a universal characterization not only of semantic or notional categories – as he and many other linguists argue⁴ – but also of grammatical categories, that is, the entities on his 'functional' level. The rest of the book is an attempt to substantiate this claim.



Examples. 1a *sheep*.—1c *can*.—1d *put*.—1e *hand*.—2a *cats*.—2b *John's*.
—2c *eats*.—3a *oxen*.—3e *frighten*.—4a *feet*.—4e *feed*

Figure 1.7

The notion of 'TMA system' and 'TMA category'

Intuitively, a 'TMA system' is the set of TMA categories of a language. This takes us back to the question of how to interpret the notion 'TMA category', and 'grammatical category' in general. As Comrie (1976, 9) points out, there is a general problem of distinguishing between what is a 'grammatical category' and what is not:

It is usual to consider the French construction *être en train de* 'to be in the process of' as a free syntactic construction that expresses progressive meaning, rather than as a grammatical category of French, although it is not clear exactly where the boundary-line would be drawn between this and the English or Spanish Progressives, which are usually considered as grammatical categories.

Since 'tenses' and 'moods' are usually thought of as morphological categories, and treated as such in traditional grammars, one might at first sight want to restrict the term 'TMA category' to inflectionally marked categories, excluding 'periphrastic' ones, i.e. categories expressed by syntactic means, e.g. auxiliaries and particles. Some linguists seem to have wanted to take such a step (see Comrie, 1976) although, as I have already suggested, this in my opinion would mean an unwanted delimitation of the field of inquiry, given the frequent cases of functional equivalence of syntactically and morphologically expressed categories across languages and even in one language – cf. e.g. the Latin 'Perfect' tenses, which are inflectional in the active voice but periphrastic in the passive.

An alternative solution would be to regard 'having morphological expression' as one of several features that characterize categories that are 'central' to the TMA system of a language. What I am suggesting is that the TMA system and in general systems of grammatical

categories are 'focused' and 'imprecise', having a centre or 'core' and a periphery, in the same way as an individual TMA category does. In addition to morphological expression, features that plausibly characterize the core categories of a TMA system might include obligatory or systematic use and (something which may be hard to distinguish from obligatoriness) lack of alternative ways of expression. The last-mentioned features would distinguish e.g. the English Progressive, which is obligatory in its typical uses, from e.g. the synonymous but optional and non-unique constructions in Swedish (*hålla på att, sitta och, vara i färd med att*), even if both the English and the Swedish constructions are periphrastic.

What I am suggesting here appears to be consonant with Comrie's view (Comrie, forthcoming) that the difference between 'grammaticalization' and 'lexicalization' 'can be understood in terms of the interaction of two parameters: that of obligatory expression, and that of morphological boundness'.

I shall return to the question of centrality in chapter 7.

The notions of tense, mood, and aspect

When defining the terms 'tense', 'mood', 'aspect', linguists usually choose a semantic point of departure. As a typical example, we may take Comrie's statement (1976, 3) that 'aspects are different ways of viewing the internal temporal constituency of a situation' or his definition (Comrie, forthcoming) of 'tense' as 'grammaticalized location in time'. Such a semantic approach is not entirely unproblematic: as we shall see later in this book, it is not at all uncommon for e.g. categories that are usually regarded as aspectual to be constrained as to temporal reference. In fact (as is also acknowledged in Comrie, forthcoming), in order to use such semantically based definitions in a constructive way, we need the additional assumption that we can determine what is basic and what is secondary in the meaning of a grammatical category.

The notion of dominance discussed above (p. 81) suggests a possible way of delimiting tense, mood, and aspect from each other. As we shall see below, although universal grammatical categories typically combine semantic parameters of temporal, aspectual or modal character, it is in several cases possible to single out one of these as dominant in the sense discussed above. Thus, the category PFV typically combines 'perfectivity' and 'past time reference' – it is clear, however, that perfectivity is subject to less variation than past time reference – and there is thus good reason to regard PFV as a basically aspectual category. In a similar way, we shall see that in spite of recent arguments to the contrary, there is something to the traditional view that future tenses are just that – tenses, rather than modal categories.

Let us now look at the individual terms of the triad, starting by giving paradigmatic examples.

As examples of a minimal pair which illustrates a tense distinction, consider the following English sentences:

(1.6) It is raining today

(1.7) It was raining yesterday

The fact that (1.7), in contradistinction to (1.6), concerns an interval in time which wholly precedes the point of speech triggers the choice of the Simple Past verb form *was* rather than the Present *is*. This illustrates a number of properties typical to tense categories:

- (i) They are expressed by the choice of one of several possible morphological forms of the finite verb or the auxiliary.
- (ii) They semantically depend on the relation between the time that 'is talked of' in the sentence and the time of the speech act – what is often referred to as 'the deictic centre'.
- (iii) They have to be expressed – the choice of tense form has to be made – whether or not there is an explicit time indicator such as an adverbial in the sentence.

Without unduly restricting the notion of tense, we cannot regard any of these conditions as necessary. If we are to choose some property as being the essential content of the notion of tense, it appears natural to choose some weaker version of (ii) – most people would agree that tenses must have something to do with time. The problem, however, is whether this can be done without weakening the condition so much that we include things that we do not want to label as tenses. To answer this question, we must first have a closer look at the notion of aspect.

Here is a minimal pair which illustrates a typical aspectual distinction:

(1.8) When I got your postcard, I was writing a letter to you

(1.8a) When I got your postcard, I wrote a letter to you

The progressive aspect *was writing* in (1.8) indicates that the process of writing a letter was in progress at the time of the arrival of the postcard, without any indication of its completion, whereas the simple verb form *wrote* in (1.8a) depicts the writing as a completed event, viewed in its totality. Aspect thus has to do with the structure of the things going on or taking place in the situation described by the sentence (cf. the definition from Comrie 1976 quoted above).

The problem is that this characterization of aspect cannot be taken in isolation from time: we can see that (1.8) and (1.8a) both concern two events or processes – the arrival of the postcard and the writing of the letter – but these have different temporal relations: in (1.8), the time of the first is included in the time of the second, in (1.8a) on the other hand, the most natural interpretation is that the first precedes

the second. So aspect clearly has something to do with time, but what is it then that makes it different from tense?

As we have seen, tenses are typically deictic categories, in that they relate time points to the moment of speech. Aspects, on the other hand, are non-deictic categories. The distinction between deictic and non-deictic categories can only be used to distinguish tenses and aspects if we do not in addition to 'absolute', i.e. deictic, tenses, admit the existence of 'relative' tenses, i.e. forms that may express temporal relations between any pair of time points, regardless of their deictic status. It is the latter ones that are difficult to keep apart from aspects. Comrie (1976, 5) suggests the following way of making the distinction:

Aspect is not concerned with relating the time of the situation to any other time-point, but rather with the internal temporal constituency of the one situation; one could state the difference as one between situation-internal time (aspect) and situation-external time (tense).

We may illustrate the subtleties necessary to apply such a rule by comparing the constructions treated by Comrie (forthcoming) as having relative tense meaning and those that are treated as having aspectual meaning. Thus, in a sentence such as (1.9) the 'interpretation of simultaneity' that the Progressive verbs obtain is said to be due to a (conversational) implicature, the basic meaning being that of imperfectivity.

(1.9) While I was washing Mary was drying

The Progressive construction in English, like the corresponding constructions in many other languages, involves a present participle. When this participle appears in a non-finite construction, such as (1.10), it is said to have 'relative present time meaning', i.e. express simultaneity as its basic meaning.

(1.10) Singing 'God save the Queen' Mary was drying the clothes

While there may be arguments for such an account, we may note that it implies that if the non-finite constructions are historically older, what was once the basic meaning of the present participle has now degraded into a conversational implicature.

Summing up, the distinction between tenses and aspects is by no means clear, although everyone knows what the typical cases are like.

Let us now have a look at the third category, 'mood'. Since mood is not well represented in English, it is not very easy to find a good illustrative example from that language. We might try the opposition between indicative and subjunctive conditional constructions:

(1.11) If you are rich, you can buy that car

(1.12) If you were rich, you could buy that car

We use the subjunctive *were* in (1.12) as an indication that the proposition expressed in the embedded sentence is known to be false (that is perhaps the prototypical case) or at least highly unlikely. Traditionally, moods are said to express the speaker's attitude to a proposition or to its truth-value. A better account for most cases of moods, however, is to say that they are a grammatical way of indicating that the proposition is embedded into a modal or non-assertive context.

It appears that languages that have mood distinctions normally use them in well defined types of subordinate clauses. Which types are marked varies from language to language, however. Thus, the subjunctive in German is primarily used to mark indirect speech and counterfactual constructions, whereas the category with the same name in French is used in e.g. final and certain kinds of temporal clauses. One generalization that can be made is that the typical opposition is between indicative and non-indicative moods, where the indicative is always the more 'real' or 'asserted' member of the opposition.

Traditionally, the imperative is included among the moods: we may note, however, that in various ways, the imperative is rather special in being used primarily in imperative sentences, which have various special syntactic and pragmatic properties. We shall have rather little to say about the imperative: we may just note that in contradistinction to other mood categories, it is apparently found in all or almost all languages, and also tends to behave surprisingly alike in them all: in particular, the imperative is almost always the morphologically least marked verb form, often identical to the verb stem.

Grammatical vs. lexical categories

One constant source of confusion in the study of TMA categories – in particular aspectual categories – is the complex interrelations between grammar and lexicon. The basic observation is that in addition to the fact that some aspectual notions are expressed by morphological means in some languages, it is also true for all languages that verbal lexemes differ in their 'aspectual potential'. It is clear, for instance, that the verbs *die* and *sleep* are quite different as regards the contexts in which they occur naturally. To take a standard illustration of this fact, *sleep* but not *die* can be used together with a durational adverbial like *for two hours*. The obvious semantic correlate of this distributional fact is that *die* is normally used of punctual events, whereas *sleep* is used of prolonged states.

One common way of making the distinction is as follows. Verb lexemes differ as to their 'Aktionsart' or 'inherent aspectual meaning'; in addition, some languages distinguish different morphological forms of the same lexeme, called 'aspects', according to the context in

which the verbs are used. As often happens, the theoretically nice distinction turns out to be rather difficult to apply in practice. To start with, we encounter the problem of separating out the 'inherent aspectual meaning' from contextual influences – after all, every occurrence of a verb is in a definite context, and there is no obvious way of determining what a 'neutral aspectual context' would be like. Also, it turns out that there is an astonishing flexibility in how individual verbs may be used. Further, whereas the distinction between 'lexical' and 'grammatical' is fairly straightforward as long as we keep to non-derived verbs, it becomes really problematic when derivational morphology enters the picture. In many languages, there are derivational processes which have effects on the aspectual properties of the verbs they operate on. By definition, derivation creates new lexemes: it would appear to follow that if we keep to the distinction between 'Aktionsart' as pertaining to the lexicon and 'aspect' as pertaining to the grammar, any derivationally expressed category would have to be lexical rather than grammatical. But let us now look at the paradigm example of an aspectual category, the perfectivity-imperfectivity distinction in Russian: it turns out that many – probably most – Russian grammarians treat perfective and imperfective verb forms in terms of pairs of lexemes rather than as sets of forms belonging to the same inflectional paradigm. (For a recent example, see *Russkaja grammatika* 1980). We would then have to conclude that Russian aspect is really Aktionsart! There are obviously other ways out: one could e.g. claim that Russian aspect is an inflectional rather than a derivational category. This is the line taken e.g. in Andersson 1972, an adherent of the 'lexical:grammatical' interpretation of the Aktionsart-aspect distinction. This is not the place to argue for one or the other solution: we shall return to the problem in chapter 3, where it will be claimed that aspect in Russian, like Slavic aspect in general, has particular properties that may well be connected with their derivational-like character.

Taxonomy of situations

As we saw in the preceding section, the question of 'inherent aspectual meaning' is intimately bound up with the classification of the events, processes, states etc. that verbs signify. This is not the place to discuss in detail the taxonomy of **situations** – to use Comrie's not entirely happy cover term for the things we just enumerated – but a few remarks about the terms we shall use are necessary.

First, an ontological remark. Strictly speaking, the taxonomy we are talking about is not one of situations but rather one of descriptions or characterizations of situations. Some extensionally-minded readers may prefer to talk of classes or sets of situations. In speaking of events, some authors distinguish between 'individual events' and 'generic

events': this is basically the same distinction as the one we are now making: we could say that we are engaged in a taxonomy of generic situations rather than one of individual situations. This is important to bear in mind, since one and the same (individual) situation may be described in different ways. For instance, we referred to *sleep* as a state: however, the situation described in (1.14) might also be truly depicted in (1.13), in which case it looks more like a dynamic event.

(1.13) John slept during class yesterday

(1.14) John violated one of the rules of his school yesterday
(viz. that of never sleeping in class)

'Dynamic situation' vs. 'state'

The most salient distinction in our taxonomy is that between 'dynamic situation' and 'state'. The intuition behind this distinction is simple: we distinguish those situation descriptions that in some way involve change or movement from those that do not. The precise delimitation is of course problematic – but we are getting used to that kind of problem. What is most notable from the point of view of TMA systems is that most languages divide up their predicate phrases in at least two types of constructions, which from the semantic point of view often correspond fairly well to a 'dynamic-stative' classification of predicates. The 'dynamic' construction type typically has a full verb as its head, whereas the 'stative' construction tends to involve nouns or adjectives in predicative function, with or without a copula functioning as the dummy head of the predicate phrase. In the grammars of some languages, this distinction shows up as one between 'verbal' and 'nominal' predicates, in others, as one between 'non-stative' and 'stative' verbs. The cut-off point between these constructions may not always occur in the same place in different languages: in English, for instance, there is a considerable number of 'stative verbs' which have somewhat special properties. In any case, the distinction between the constructions tends to be of considerable importance for TMA categories, in particular for aspectual categories, in that those categories tend to be less developed or wholly neutralized in stative contexts. Another somewhat different example is Bulgarian, where the distinction between Imperfect and Aorist is neutralized in copula constructions. A not uncommon case is also for aspectual categories to develop special functions in stative contexts. We do not have to go further than English to find an example of this: the Progressive is found in its primary function (that of signalling an 'on-going process') only with non-stative verbs; with statives it is only used with special interpretations.

There are also examples of tense categories being restricted to either stative or dynamic contexts: for example in Beja, past time reference is marked only in dynamic constructions. In Afrikaans, only

the copula has a morphologically marked past tense: in all other cases, a periphrastic construction is used.

In the following, we shall often use STATIVE or DYNAMIC (or the subscripts 's' and 'd') in category labels. (For the use of capitals for words in the text, see the preface and p. 29.) This should not be interpreted to mean that they are to be taken to be TMA categories in themselves; rather, they are used to indicate that a category is restricted to one type of contexts. It is important to bear in mind the impreciseness of the distinction: as was noted above, the exact borderline between 'stative' and 'dynamic' is bound to vary.

Boundedness

Within dynamic situation descriptions, the most important subclassification is that between bounded and unbounded situation descriptions. I have discussed this concept at some length in earlier works, in particular Dahl 1981, and shall not repeat all the details of that discussion. In the following, the term 'bounded' should be understood in the following way: A class of situations or a characterization of a situation is bounded if and only if it is an essential condition on the members of the class or an essential part of the characterization that a certain limit or end-state is attained. Thus, 'run five kilometres' specifies a bounded class of situations, viz. those in which a person runs at least five kilometres, and no process where this limit is not reached could be truly characterized by the phrase in question. 'Run fast', on the other hand, is an unbounded characterization, since no limit is involved. For discussion of various borderline cases and other problems, see Dahl 1981.⁵

Reichenbach's theory of tense

Reichenbach 1947 contains a section entitled 'The tenses of verbs'. We shall adopt the terminology used there, since it is quite useful to work with and is in addition by now fairly widely known among linguists. This, of course, does not imply that we necessarily swallow Reichenbach's whole theory of tense. In particular, we shall see that Reichenbach's system is essentially Boolean, which means that it has the restricted logical power of such systems.

Reichenbach claims that there are at most three points in time which are relevant to the choice of tense in any given sentence:

- S – the point of speech;
- E – the point of the event;
- R – the point of reference.

We can illustrate these three points if we consider an utterance such as (1.15):

(1.15) Peter had gone away

S is the time when (1.15) is uttered. E is the time when Peter went away, and R is some definite time-point between S and R provided by the context. The Past Perfect is one of the few cases in English where S, E and R are all different – most often, two of them coincide. One of the main points of Reichenbach's theory is that he is able to distinguish the Perfect tenses from the simple Past by assuming that in the former, E and R are different (in the Present Perfect R coincides with S). We shall discuss this further in chapter 5.

The reason that Reichenbach's system is Boolean is that all tenses are supposed to be describable in terms of precedence relations between a finite set of points in time and thus, the characterization of a tense can be reduced to the conjunction of a set of propositions, chosen from among a finite set. (Comrie, forthcoming) suggests a revision of Reichenbach's system in which, among other things, there may be an unlimited set of reference points R_1, R_2 etc.)

Reichenbach's scheme works best for cases involving single, punctual actions: his accounts of the Progressive in English and the Aorist in Turkish are not convincing. That is partly due to the fact that the scheme basically only accounts for temporal reference and is not well suited for describing aspectual notions. However, there is at least one way in which the scheme can be made more powerful. Consider the following sentence:

(1.16) When I arrived, Peter had tried to phone me twice during the preceding week

R here must be the time when I arrived, and E – or rather the E's – are the time-points when Peter tried to phone me. But we see that there is nothing in Reichenbach's scheme that corresponds to the time referred to by *during the preceding week*. The function of that phrase could be said to be to specify a **temporal frame** for the E points, that is, a time period within which they are located. Let us call the temporal frame F. In fact, although complex cases like (1.16), where S, E, R and F are all distinct, are rather infrequent and perhaps a little far-fetched, it is quite common that we need to postulate an F in order to give an adequate account of the interpretation of a sentence. Consider e.g.:

(1.17) Last year (=F), the fall semester began on 29 August (=E)

We shall later argue that the concept of 'temporal frame' is necessary to account for the use of various TMA categories in our material.

There are in fact several ways in which a temporal frame can be determined. As in (1.17), it may be introduced by an explicit time

adverbial in the sentence. It may also have been introduced explicitly or implicitly in the earlier context, as in:

(1.18) I had a nice time in London. I went to Madame Tussaud's twice (i.e. during the time I was in London)

Finally, but not least importantly, we may know, for various reasons, that the type of event talked about could not take place except within a certain period of time. For instance, a sentence such as (1.19) will as its F have the (adult?) life-time of the speaker's grandfather.

(1.19) My grandfather went to his dentist twice every year

F here gives the interval over which the quantifier *every* ranges. We also need to invoke F in order to explain why English avoids the Present Perfect in sentences like (1.19), if the person referred to is no longer alive.

Cross-linguistic generalizations about TMA categories

The nature of cross-linguistic generalizations

The ultimate aim of a typologically oriented study of TMA systems is to determine in what ways such systems in human languages are similar to each other and in what ways they may differ. For these questions to be meaningful, there must be some limit to variation among languages: this is probably a relatively uncontroversial assumption today, but only thirty years ago the prevailing dogma in at least some of the most influential schools of linguistics was exactly the denial of it. The search for cross-linguistic generalizations is often seen as a quest for 'language universals', i.e. properties that are common to all human languages. In actual practice, it is quite seldom that absolute, non-definitional universals are identified in data-oriented work – properties that can truly be said to be non-vacuously manifested in **all** human languages tend to be of a rather abstract character and often only extremely indirectly testable (in the best case!). More commonly, the universals found by typologically oriented linguists are of a weaker kind – implicational or statistical. The claims I want to make on the basis of the investigation reported here are no exception in this regard. I will not claim that all languages use the same TMA categories but only that the overwhelming majority of all categories found in the TMA systems of the world's languages are chosen from a restricted set of category types. I do not find that I have evidence enough to say that the set of categories that I am postulating totally exhausts what will be found in human TMA systems. It appears to me that the view that we are born with a universal grammar in our heads that tells us exactly what is possible

and what is not is misguided. To choose a non-linguistic comparison, we are clearly predisposed to walk on our feet: yet, to say that our genetic equipment makes it positively impossible that there could be a culture where the primary mode of locomotion is walking on one's hands seems a bit too strong – or rather, it is hard to imagine that physiologists would spend a lot of time discussing whether such a culture could exist or not, given sufficiently extreme situational preconditions: the question is simply not very interesting.

In recent years, there has been a clear development away from such a simplistic idea of a universal grammar. For example, Chomsky and others have argued that the innate linguistic competence specifies a number of 'parameters' and the default values of these parameters that will be chosen by the language learner in the absence of evidence to the contrary. Yet Chomsky still favours the view that universal grammar specifies the limits of what is possible in human languages. I would prefer a view where our innate competence determines (rather than 'specifies') what is expected in a language but where there are no sharp limits to the possible deviations from the default states: that is, it is not excluded that the speakers of some languages 'walk on their hands' in a metaphorical sense. We do not yet understand very well how it comes about that some languages go to extremes in developing arcane subcomponents of their grammars and phonologies: yet this phenomenon is found all over the place – think of the click systems of the Khoisan languages, the politeness system of Japanese or, say, the extraordinarily rich tense-aspect systems of some Bantu languages. Maybe Sapir's term 'drift' may be apt here: it appears that some languages 'cultivate' some phenomena, and we cannot tell how far such cultivation may go.

Some linguists object to the use of the term 'universal' for phenomena that are not positively manifested in every human language; yet such a usage has a solid tradition. It may be noted, for instance, that in the Jakobson-Fant-Halle theory of universal phonetic features, there was no claim that every member of the set of universal contrasts should be exploited in every language. The terminological question is of course of restricted importance, but the reluctance against calling non-absolute universals 'universals' is often connected with the implicit or explicit view that nothing short of a perfect absolute universal is of interest to general linguistic theory. I think this view emanates from the general fixation on universals that has characterized linguistics over the last thirty years. It may indeed be time for a terminological reform: I shall in the rest of the book try to use 'cross-linguistic' instead of 'universal' wherever it is possible, in order to divert the attention from the question of absolute universality. What it is essential to retain from the universalist tradition is the refutation of the relativist view that 'every language must be described in its own terms', and that there is a restricted set of categories which

will be sufficient to account for the basic traits of any human language, in the normal case. Although this looks like a rather hedged statement, it should be emphasized again that it is no weaker than most general claims in science.

The basic units of the cross-linguistic level of description

Assumptions about linguistic universals during the latest three decades have been heavily influenced by the example set by Jakobson and Halle in their theory of universal phonology, where the basic units consisted of a small, assumedly universal set of binary features. It has been widely assumed, explicitly or implicitly, that this kind of model is valid also outside of phonology. Applied to TMA systems, this would mean that the final result of our undertaking would be the establishment of a similar set of underlying (in our case apparently semantic) primitive features, linguistic quarks as it were, out of which all TMA systems would be built up.

The approach I shall take in this work is different. Continuing the metaphor from physics, I shall suggest that the most salient 'universals', or better, basic units of the general theory of TMA systems are rather atoms than elementary particles, i.e. categories rather than features. More concretely speaking, this means that I think of a language-specific TMA category like, say, the English Perfect, as the realization of a cross-linguistic category – or better, category type – PERFECT, rather than as the realization of a set of features, say, /+X, -Y, +Z/.

To explain why I have chosen this kind of approach, it is necessary to give more details about the assumptions I make about cross-linguistic TMA categories. Connecting to the discussion above of impreciseness and focusing, what I want to claim here is that the main criterion for identifying TMA categories cross-linguistically is by their foci or prototypical uses, and that languages vary essentially in two respects: (i) which categories they choose out of the set of cross-linguistic categories, (ii) how they reduce the impreciseness that these categories have in choosing among the possible secondary or non-focal uses they have. The universality of foci of categories is of course no new idea: it was put forward for the first time, as far as I know, in Berlin & Kay's classical study of colour terms (1969). To my knowledge, it has not been applied earlier to a semantic field with such direct relevance for grammar as TMA, at least not in a large-scale data-oriented investigation.

The foci and extensions of categories may be seen as points and regions respectively in a multi-dimensional 'conceptual space'. In a way, the dimensions that make up the space will take the place of the ultimate building-blocks – 'quarks' – in the theory. An immediate question is then what these dimensions are, if they are universal and

how many there are of them. It is reasonable to assume that the basic structure of 'conceptual space' is innate – even an empiricist like Quine concedes that it is hard to see how learning could otherwise take place (1960, 83). (Adherents of conceptual relativism among linguists mostly miss this point, though: they talk about the arbitrariness of how language 'cuts up' reality without realizing that this way of speaking presupposes that there is something to be cut up in the first place.) It may well be difficult, however, to delimit the set of dimensions that are relevant to a given linguistic category: one of the things that become clear when you are trying to describe the interpretations of linguistic expressions in terms of prototypes is that the number of properties or features that go into the characterization of a prototype is not necessarily well determined. This is one of the reasons that I do not think that one can re-interpret Jakobson-style features as dimensions of the conceptual space.

In any case, the result is that we must recognize two levels in general theory: the level of (universal) categories or category types and the level of dimensions of conceptual space. The practical importance of this distinction becomes clear when we see that the same traditional label may play a role on both levels, with subtle but crucial differences. Thus, 'past' may correspond both to a cross-linguistic TMA category – for which we shall use an upper case denomination (PAST) – and a value of an underlying dimension – the names of which we shall put in single quotes. The latter in its turn may play a role both in the category with the same name but also, crucially, as a feature of the category PERFECTIVE.

In addition, labels like 'past' will have to be used for language-specific categories, such as for example the Past (tense) in Russian. For these, we shall, following Comrie 1976, use initial capitalization. Notice, however, that where we have introduced three levels of description, Comrie operates with two levels only: universal semantic categories (in his book denoted by lower case letters) and language-specific grammatical categories.

There are further reasons for distinguishing between semantic and grammatical categories on the level of cross-linguistic description. When one starts to classify language-specific categories in terms of which cross-linguistic category types they instantiate, it quickly becomes clear that this classification is strongly correlated with the ways in which the language-specific categories are expressed. To start with, there is in general wide-reaching consistency as to which member of a given opposition is the marked one (in the sense of 'marked' defined above). Furthermore, there are also significant differences between cross-linguistic category types as to what we shall call **marking type**, i.e. whether a category is marked by morphological means or periphrastically, e.g. by auxiliaries.

The consequence of this is that it is possible to characterize the

typical morphosyntactic behaviour of a given cross-linguistic TMA category in a way which is similar to the characterization of its semantics: it becomes possible to say that a cross-linguistic TMA category is associated with a cluster of semantic and morphosyntactic features, although no subset of this cluster can be singled out as providing necessary-and-sufficient conditions defining the category. We may find an analogue to such a claim about a grammatical category in Keenan 1976, where it is proposed that the notion of 'Subject' in universal grammar is best defined in terms of a set of semantic and syntactic properties which are characteristic of typical subjects in any language.

2

The investigation

As was noted in chapter 1, one may divide the empirical work that lies behind this book into two parts: first, the creation of a data base aiming at as complete a coverage as possible of the TMA systems of a representative sample of the world's languages; second, an analysis of this material which was to test the hypothesis that TMA systems may be described in terms of a limited set of cross-linguistic category types. One half of the chapter will be devoted to each of these two parts of the investigation.

Creating the data base

The first part of the investigation – the creation of the data base – involved at least the following stages, not necessarily separate in time:

- (1) Planning the investigation and constructing the questionnaire.
- (2) Finding informants and collecting data.
- (3) Analysing individual questionnaires and storing data on a computer.
- (4) Writing reports on individual TMA systems.

The account of these stages which follows will be problem-oriented, that is, I shall try to discuss as much as possible general methodological questions that are bound to arise in investigations of this kind. I feel that this will make the account more interesting to the general reader than it would be if I put the emphasis on the details of what we actually did.

Choice of data

Typological work in general may be based on two types of data: (i) 'primary data' obtained mainly by working with informants or directly observing language use, (ii) 'secondary data' obtained from extant descriptions of languages. Originally, we wanted to make use of both kinds of sources in the investigation. However, in the actual work, we

came to pay less attention to secondary data than planned. From the beginning, we planned to compile a relatively large number of sketches of TMA systems of various languages based on grammars and descriptions. It turned out, however, that this was less rewarding and more time-consuming than we had thought. It is often extremely difficult to draw any definite conclusions from such descriptions, due to the vagueness of the terminology and, in many cases, difficulties in interpreting the examples given. Also, it is not uncommon to find discrepancies between a grammar and actual use, in that forms and constructions are either completely neglected or mislabelled.

It was therefore decided to try and base the work mainly on primary data, collected with the help of a questionnaire, called the 'TMA questionnaire'. The main part of this questionnaire consists of a number of sentences and short connected texts in English together with indications of the contexts the sentences or texts are assumed to be uttered in. These sentences and texts were then translated into the languages to be investigated by native informants. The questionnaire is reproduced in appendix 1 and is discussed in more detail below.

Choice of languages

In contrast to the situation in the behavioural sciences, where questions of sampling techniques – like other methodological problems – have been paid constant attention, linguists have not in general worried very much about how they choose a sample from a population. Neither typologists nor students of tense and aspect are exceptions in this regard. The first paper to seriously discuss sampling in typology is Bell 1978.

The main sampling problem for the language typologist is how to guarantee that the set of languages from which he is making his generalizations is not in any way biased. As Bell points out, even linguists who try to extend their investigations to non-European languages usually end up with a genetic – Indo-European – or areal – European – bias or both in their material. Even among what a European linguist would normally regard as 'exotic' languages there are enormous differences as to availability of informants and descriptions. It is easy to think that you have done your duty when you have quoted token examples from Arabic, Chinese, Turkish and Swahili. To fully overcome the bias problem would demand superhuman abilities. One fact not mentioned by Bell should be contemplated in this connection. The concept of 'areal' bias relies on the assumption that languages that are in close contact influence each other. The word 'areal' is misleading, however, in that one is led to think of languages as being spoken in one definite place. In our times, though, the most important interlinguistic influence is exerted by languages with world-wide distribution like English and other 'imperialist' languages

upon the languages of less powerful peoples, largely independently of the geographical areas where the latter are spoken. The problem is that there is a direct correlation between the degree of accessibility of a language to a student of typology and the extent to which this language has been exposed to European influence. To take an illustrative example, it is quite probable that a language such as Guarani, which has official status in Paraguay, has been more influenced by Spanish than some language only spoken in an isolated valley in the Andes. At the same time, the chance that the former will get into someone's typological sample is very much larger than the chance that the latter will, for obvious reasons. We simply cannot tell how much this actually biases our samples. In our material, the European influence on the vocabulary of 'colonized' languages is readily observable: even if it may be safely assumed that grammar is less easily influenced than the lexicon, seeing how many languages have e.g. borrowed a word like *cup* from European languages is a healthy reminder of what we have said here.

There are several other biasing factors of a similar kind. Languages that are accessible to the typologist tend to (i) have a large number of speakers, (ii) be spoken in industrialized societies with a high literacy rate, (iii) have long traditions as written languages. It is not possible a priori to exclude any of these factors as having influence on the phenomena that one wants to study. Again, however, getting around them seems to demand a superhuman effort.

An important fact to bear in mind when discussing the choice of languages for a typological investigation is that it is not necessarily the case that it is desirable or even possible to make a choice of languages once and for all. Given the great differences in availability of information, looking for an ideal sample will very soon lead the investigator to rejecting information about close-at-hand languages. At the same time, there are a number of reasons why it would be rather stupid to do so in an investigation of the present kind. Having access to first-hand data about some well studied language makes it possible, among other things, to compare one's own data with extant descriptions of this language, thus enhancing the possibilities of evaluating both. Also, a comparison of several closely related languages may well throw light on the ways in which almost identical grammatical systems may differ in details, and suggest how diachronic processes may influence the grammar. A further consideration is that having one questionnaire from each of a number of closely related languages at least partly serves the same purpose as several questionnaires from the same language: in both cases, having several instances of one and the same phenomenon makes it easier to exclude the possibility of an error of any of the types discussed below.

There is therefore good reason for making a distinction between the total set of languages included in one's material and the sample to be

used whenever statistical generalizations are called for. In practice, what happens – or at least, what happened in our case – is that one takes whatever chances one gets of obtaining a completed questionnaire for a language, regardless of its genetic and geographical affiliations. However, in order to get at least a minimal coverage of the major geographical areas of the world, extra efforts have to be put in for the remoter ones. The methods we used to collect questionnaires therefore had to be quite varied, ranging from asking colleagues and friends to putting up announcements at Swedish courses for foreign students and systematically searching directories of institutions engaged in linguistic research in other countries.

With very few exceptions, the questionnaires we received were of acceptable quality (as far as we could judge). Those that were not have not been included in the final material.

In the majority of cases, we had to be content with one informant from each language: sometimes, however, we were lucky enough to obtain two or more completed questionnaires. For a number of languages, several persons collaborated on one questionnaire. In the ideal case, of course, one would like to have the judgements of a large number of speakers, but this could hardly be realized for the number of languages we were dealing with. As mentioned above, it is possible to compensate for this to some extent by having several questionnaires for closely related languages.

Table 2.1 The languages in the sample by genetic groups

	Code
AFRO-ASIATIC	
<i>Cushitic (AA1)</i>	
Beja	BJ
Oromo (Galla)	OR
<i>Semitic (AA2)</i>	
Amharic	AM
Arabic (Modern Standard)	AB
Arabic (Tunisian)	AR
Hebrew	HE
Maltese	ML
Tigrinya	TG
ALTAIC	
<i>Turkic (AC1)</i>	
Azerbaijani	AZ
Turkish	TU

Table 2.1 continued

<i>Korean-Japanese (AC2)</i>		
Japanese		JA
ANDEAN-EQUATORIAL		
<i>Andean-Quechumaran (AE1)</i>		
Quechua		QU
<i>Tupi (AE2)</i>		
Guarani		GA
AUSTRALIAN (AN)		
Alawa		AW
Bandjalang		BA
AUSTRONESIAN		
<i>West Indonesian (AU1)</i>		
Cebuano		CE
Indonesian		IN
Javanese		JV
Sundanese		SD
<i>South Celebes (AU2)</i>		
Bugis Makassar		BG
<i>Polynesian (AU3)</i>		
Hawaiian		HA
Maori		MR
CAUCASIAN		
<i>Kartvelian (CN)</i>		
Georgian		GO
DRAVIDIAN (DR)		
Tamil		TA
ESKIMO-ALEUT (EA)		
Greenlandic Eskimo		GL

Table 2.1 continued

INDO-EUROPEAN		
<i>Germanic (IE1)</i>		
Afrikaans		AF
English		EN
Fitzroy Crossing Kriol		FC
German		GE
Swedish		SW
<i>Indo-Iranian (IE2)</i>		
Bengali		BE
Hindi/Urdu		HI
Kurdish		KU
Punjabi		PU
Persian		PE
<i>Greek (IE3)</i>		
Greek (Modern)		GR
<i>Romance (IE4)</i>		
Catalan		CA
French		FR
Italian		IT
Latin		LN
Limouzi		LI
Romanian		RO
Spanish		SP
Portuguese		PG
<i>Slavic (IE5)</i>		
Bulgarian		BU
Czech		CZ
Polish		PO
Russian		RU
IROQUOIS (IQ)		
Oneida		ON
Seneca		SE
KAM-TAI (KT)		
Thai (Siamese)		TH

Table 2.1 continued

MON-KHMER (MK)	
Kammu	KM
NIGER-CONGO	
<i>Bantu (NC1)</i>	
Kikuyu	KY
Sotho	SO
Zulu	ZU
<i>Gur (NC2)</i>	
Karaboro	KB
<i>Kwa (NC3)</i>	
Işekiri	IS
Yoruba	YO
<i>Volta-Comoe (NC4)</i>	
Akan	AK
<i>West Atlantic (NC5)</i>	
Wolof	WO
SINO-TIBETAN (ST)	
Chinese (Mandarin)	CH
URALIC	
<i>Finno-Ugric (UC)</i>	
Estonian	ES
Finnish	FI
Hungarian	HU

In table 2.1, the languages that were included in the investigation are listed, sorted by genetic classification. The total number of languages is 64. The genetic group which is best represented is, for natural reasons, Indo-European with 21 representatives, i.e. about a third of the total sample. Relatively well represented are also e.g. the Semitic and the Niger-Congo languages.

Looking at the sample from an areal point of view, we can see that

all the major continents are represented by at least two languages. There are, though, some areas that are not represented at all, including two where according to current estimates in all up to 3,000 languages are spoken, viz. New Guinea and the NE part of South America. Admittedly, these areas belong to the least accessible and, in general, least extensively studied in the world – even the estimate of the number of languages is far from certain. Thus, the lacunae in the sample are irritating but excusable.

In order to correct possible biases in the material, I decided to choose a smaller sample of languages, which could be used whenever statistical generalizations about frequencies among the world's languages were called for. To this end, I divided the languages in the sample into 11 'areal-genetic' groups: this classification is more or less in accordance with earlier proposals of the same kind (as found e.g. in Bell's paper). From each areal-genetic group I picked out no more than two languages at random, but seeing to it that choices of closely related languages were avoided, if possible. In this way, I hope to have obtained a maximally representative sample, given the limitations of the material. The result is seen in table 2.2.

Table 2.2 Languages in the small sample

Australian	Alawa, Bandjalang
Austro-Tai	Bugis Makassar, Sundanese
Austro-Asiatic	Kammu
Sino-Tibetan	Chinese (Mandarin)
Dravidian	Tamil
Boreo-Oriental	Azerbaijani, Hungarian
Indo-European	French, Kurdish
Caucasian	Georgian
African	Beja, Kikuyu
North Amer-Indian	Greenlandic Eskimo, Seneca
South Amer-Indian	Guarani, Quechua

Choice of informants and investigators

Two roles in the filling-out of the questionnaire were distinguished: the 'informant' and the 'investigator'. In some cases, one person could perform both functions. This, naturally, was the case for languages native to members of the project. The same held for the cases when an external informant was used who was also a trained linguist or philologist. When the investigator was different from the informant, the former could be either an external 'expert' on the language or when no such person could be found, a member of the project. In the

latter cases, we experienced all the problems of the classical 'field linguist' who describes a language totally unknown to him. Naturally, these were the occasions when we had to rely on written grammars to a large extent.

As for informants, the aim was to use only native speakers. (A couple of questionnaires that were filled out by non-native speakers and where we could compare the result with the output of native speakers confirmed that this was a wise requirement. TMA categories notoriously belong to the most difficult things to master in a foreign language.) In some cases, of course, it was not possible to establish exactly what the competence of an informant was: living for a long time in a foreign country often has a negative effect on your proficiency even in a language that you used to know as a native. This, however, is an unavoidable difficulty in a large-scale investigation of this kind: even if you are lucky enough to be able to make real field studies, you may encounter the same problem with languages that are extinction-bound, where the informants that you can get hold of are often old and may not have actively used the language for a long time. At least one questionnaire had to be rejected since the informant turned out not to know even simple lexical items in his purported native language.

One exception to the native speaker requirement was made when it was decided to include Latin in the sample, the consideration being that in the absence of true native speakers of this language, one might still be able to get translations that were correct enough to merit inclusion in the material, taken of course with a grain of salt.

The TMA questionnaire (TMAQ)

The TMA questionnaire (TMAQ) has existed in at least three versions, of which the last two are relatively similar. The first version was tested on a smaller number of languages. It turned out to have various shortcomings, and in constructing the later versions we tried to remedy those. With one or two exceptions, the third version has been applied to all the languages in our sample.

In its third version, the questionnaire consists of three parts (parts A–C). Parts A and B contain sentences and texts respectively to be translated into the language under investigation. In part C, the investigator was supposed to give a list of all relevant TMA categories in the language and provide sample paradigms of lexical items. We also had a part D, which consisted of charts where the collocational possibilities of different TMA categories and temporal and other adverbs were to be listed. We soon found, however, that this was too ambitious a task, and part D was discarded altogether.

Figure 2.1 shows the format used in part A. In the box 'CONTEXT', the linguistic or extralinguistic context of the utterance was

No. 1	ANALYSIS
CONTEXT	
Standing in front of a house	
Sentence to be translated (omit material with parentheses)	
The house BE BIG	
Same translation as sentence No. <input type="text"/>	TRANSLATION:
	1

Figure 2.1

indicated (see discussion below). In the box 'Sentence to be translated', the sentence was given in English. The nucleus of the predicate (the verb or the verb + following adjective) was capitalized and given in the infinitive. The point of this was to minimize influence from English when translating (i.e., we wanted to avoid literal translations of English TMA categories).

The above means that the information necessary to choose the correct TMA category in the translation would have to be deduced from the sentence itself together with its context. In some cases, the interpretation of a sentence was made more precise by phrases or words added in parentheses which were not to be translated. (This turned out not to be quite unproblematic, see below.)

The informant/investigator was instructed to write the translation of the sentence in the box labelled 'TRANSLATION', using the most natural way of expressing the English sentence under the given interpretation. If there were several equally natural translations, the instructions were to indicate them all. However, if there were several possible translations, and one of them was of a clearly periphrastic character, i.e. was syntactically considerably more complex than the others, it was to be omitted. This was done in order to avoid proliferation of optional alternatives. In actual practice, however, this instruction was not always carried out.

In part B, which consists of connected texts, contexts are only given once for each text. The format is as shown in figure 2.2.

The first stage of the investigation, then, was for the informant to translate the sentences in parts A and B into his/her own language. Next step was for the investigator to provide every verb¹ in the questionnaire with a code identifying its TMA category (this was done

No. B6

CONTEXT		
I'll tell you what happens to me sometimes, when I am walking in the forest.		
Same translation as text No. <input type="text"/>		
Text to be translated	TRANSLATION	ANALYSIS
a. I SEE a snake		B6a
b. I TAKE a stone and THROW at the snake		B6b

Figure 2.2

in the ANALYSIS box). To this end, it was necessary to give a list of the TMA categories of the language in question together with an alphanumeric code for each category. This list was to be given in part C, where also illustrative paradigms were to be given (of the verbs *die*, *work*, *read*, and *sleep*).

This work, of course, was not without its problems. Since, in many cases, we had to rely on the help of external investigators, we had to give relatively straightforward principles for choosing codes.

It is to be noted that at this stage of the investigation, no attempt was made to arrive at universally valid semantic classifications of categories. The minimal requirement that we put on the coding was that forms that differed morphologically should be distinguished. The fact that we chose e.g. numbers beginning in 1 for forms with present time reference was purely a matter of convenience.

In fact, we tried to make the analysis as 'superficial' as possible in the sense that whenever there was a choice between criteria of meaning and criteria of expression for identifying categories, we chose the latter. That is, if two forms seemed to have the same semantics although they were different in their expression, we preferred to keep them apart. The motivation for this was that having identified two seemingly semantically identical categories, we might in fact hide differences that would be apparent given a deeper

analysis. One example would be languages like German, where different auxiliaries (*haben* and *sein*) are used to form the perfect of different verbs. Identifying these two kinds of perfect would mean hiding the fact that the choice between them is in fact dependent on semantic criteria.

One consequence of the 'superficiality' principle that may seem self-evident but which it may be wise to point out is that it should never be the case that two categories are expressed the same way and differ in meaning only. That is, we excluded analyses like the ones found in Latin-inspired grammars of English where English was said to have five or six cases, or to construct an example from the TMA area, one where English verbs would be said to have homonymous forms called 'Imperfective Past' and 'Perfective Past'.

Problems in constructing the questionnaire

One problem that very soon faces the constructor of a typological questionnaire is what we can call the 'combinatorial explosion'. We naturally want to cover the area of study as well as possible; we would thus like to have at least one example of each relevant combination of parameter values. However, the number of such combinations grows very quickly, as the number of parameters is increased, and will soon exceed the limits that practical considerations put on the size of a questionnaire. For instance, the (approx.) 200 sentences in our questionnaire fill about four pages of typewritten text (if each sentence takes up one line); yet, 200 is less than 2^8 or 3^5 , which means that we may investigate no more than seven binary or four ternary parameters. Already a set of standard syntactic parameters such as 'declarative/interrogative', 'active/passive', 'main clause/dependent clause', 'affirmative/negative' takes up four of our seven binary choices, and it is easy to see that the rest could be filled out very quickly. No wonder then that it was not possible in our investigation to look at more complex cases or marked constructions in the detail that would have been desirable. In particular, it would have been most interesting to investigate more closely e.g. embedded clauses, negated sentences, and passives, since it is well known that TMA categories behave in peculiar ways in these contexts.

In addition to the combinatorial explosion, there are additional reasons why it is hard to study e.g. passive constructions with a methodology of this type. In some languages, passives may be very marginal or even not occur at all; other languages may have several constructions that could be labelled passive, with more or less subtle differences in conditions of use. This means that by simply giving an English sentence in the passive you cannot guarantee that what comes out in the translation is the passive construction you are looking for, if it is a passive at all.

Choosing the sentences also turned out to be connected with various problems. Sentences containing culture-bound items may cause problems when translated: we had to skip several examples from the first versions (such as *The boy drank a glass of milk*). On the other hand, the sentences should not be too trivial or uniform; otherwise the translators will get bored very quickly. We tried to circumvent the dilemma by giving the informants the possibility of substituting problematic expressions (e.g. 'desert' for 'forest'). Still, we got some complaints from informants and investigators about cultural biases²) in the questionnaire. After this experience, I would be prepared to announce an award to anyone who is able to come up with, say, two examples of completely non-culture bound transitive sentences with a human subject and a predicate phrase denoting an activity.

Sources of errors

There are, naturally, a large number of ways in which things may go wrong when collecting typological data by the questionnaire method (as with any other method of data collection, of course). We shall now consider some different types of possible errors, how they can be detected and how they can bias the results of the investigation.

What we are looking for are underlying patterns in the structuring of TMA systems. Various kinds of errors will create 'noise' which may prevent us from perceiving those patterns. This may happen in basically two ways: either the noise simply hides the structures that are there, making things look more chaotic than they actually are, or – which is more serious – it creates apparent order, displaying false patterns. When discussing types of error, it is of some importance to consider how they will most probably behave in this regard.

The first step in the process is, as we have seen, that the informant translates the sentence in question into his own language. There are a number of prerequisites for the success of this operation. In formulating a typological questionnaire of the type under discussion, the linguist has certain intentions: for every sentence, he has a given interpretation in mind. The problem of translation therefore is a two-way one: first, for the linguist to communicate his intention to the informant, then, for the informant to find the best way of expressing this intention in his language. (If something goes wrong, it is not necessarily the informant's fault!) Let us consider the linguist's problem first. As anyone who has tried this task will have realized, communicating exactly what you want is not always easy. Some sentences turned out to be systematically misunderstood by the informants: even if a number of such cases were discovered at an early stage and were not included in the final version of the questionnaire, a couple seem to have remained. Happily, they can be rather easily

discerned, at least in the more blatant cases: if a sentence where you were expecting a past tense is systematically translated as a future into the most different languages, one starts suspecting that something is wrong. More dangerous, of course, are those cases which are misunderstood only part of the time. This takes us to the question of the informant's competence. We have already discussed the condition that the informant be a native speaker of the target language: even if the demands need not be as high for the source language, i.e. English, a minimal level of proficiency is of course needed. This is something which may be hard to check in every case, in particular when questionnaires have been sent to us from other places. Clearly, if the level of comprehension is **too** low, this will result in massive failure on the part of the informant, something which will be immediately obvious. More subtle misunderstandings may be harder to detect, however.

It should be pointed out that there is sometimes the possibility of the investigator translating the sentences into an intermediary language known to the informant. Since this introduces a further step, it naturally increases the risk for errors. We do not know if our corresponding investigators have had to use this method and to what extent.

Even a competent informant may of course fail to translate a sentence in the desired way. Since the task requires attention and concentration, a less well motivated or tired informant is likely to translate the sentences inexactly or sloppily. The possibility that the informant simply becomes exhausted is not to be neglected when dealing with a questionnaire of this length! In particular, since the informant has to figure out the exact interpretation via the indicated contexts, someone who pays insufficient attention to these indications is bound to give bad translations.

Also, even a person with very good command of the languages involved may not be very good at translating from one of them to the other. Teachers of foreign languages will have observed that pupils who are asked to translate something may produce ungrammatical output even in their own language. In general, it may be suspected that translations are often rather unrepresentative of the translator's normal language, in particular if he is not a professional. Together with a group of students at the Department of Linguistics in Göteborg, I made a comparison of different methods for studying how speakers of a language choose between alternative constructions or forms (we looked at the choice between reflexive and non-reflexive pronouns in Swedish). The general impression (which was not based on any statistical analysis of the material) was that the method of letting speakers translate sentences from a foreign language into their own was the least reliable of the methods we tested, that is, it showed the lowest correlation with the others.

If what we have said here is true, one may ask how anyone could ever be so stupid as to choose translations as a basis for an investigation of language use. The simple answer is that it is the only realistic method for large-scale data collection in typologically oriented linguistic research. We simply have to accept that it is unreliable and try to use the data with the necessary care. For instance, one should avoid drawing conclusions from single cases – if you are looking at the translation of one sentence into one language, the probability that that particular translation is inadequate is too large to make it possible to infer anything from it at all. Patterns that are manifested in a large set of sentences in several languages, on the other hand, are much more trustworthy, in particular if they are corroborated by previous descriptions of those languages.

A linguist working with an informant and a questionnaire of course has the possibility of checking the informant's translations by putting questions to him. Naturally, we have tried to use this possibility whenever we have had direct contact with the informant. It is harder to do such things, however, when you obtain a questionnaire by mail from a far-away place.

As can be seen from appendix 1, the sentences in the questionnaire were at least partly ordered in such a way that similar sentences or one and the same sentence put in different contexts were grouped together. I now think that it might have been better to randomize the order. It is unclear what effect the present order has: some persons who have filled out the questionnaire have pointed out to me that the result may be that one tries to differentiate contiguous sentences whenever possible. This probably leads to an exaggerated use of 'marked' constructions, which may mislead the investigator but which may also, in some cases, make it easier for him/her to discern the patterns that exist.

The computer analysis

The computer analysis was not part of the original project plan, but was decided upon when we realized the problems in coping manually with the data we had obtained. What was fed into the computer were the already analysed data – that is, the category codes discussed above. It is possible that we would also have stored the informants' translations of the questionnaire sentences, if we had planned the investigation on a computerized base from the start. Now, we decided that the possible gains we could get from this did not make up for the amount of extra work it would entail.

In the first stage, we stored each questionnaire as a separate file, where each sentence in the questionnaire corresponded to one numbered line. This was done using the GUTS system on the IBM

computer at the university computer centre in Göteborg. Later, it was found that in order to be able to analyse the data in a more sophisticated fashion, we needed a data base system. For this purpose, the files were transferred via a tape to the DEC10 computer at QZ Computer Centre in Stockholm and converted so as to be analysable by the 1022 data base system.³

In data base systems, data are usually stored in the form of 'records' which are collections of 'attributes'. For instance, if the data base contains information about persons, each record might correspond to one person and the attributes might be 'first name', 'last name', 'date of birth' etc. We could display a record as follows:

FIRST NAME	LAST NAME	DATE OF BIRTH
John	Smith	06-01-1950

In our primary data base, a record corresponded to one translation variant of one verb in one questionnaire. The following illustrates what such a record may look like:

LANGUAGE	SENTENCE	VERB	VARIANT	CATEGORY	COMMENT
SW	009	1	1	2	—

This means that the first verb of sentence 9 in the questionnaire was translated into Swedish (SW) using the Past tense. The number '2' under CATEGORY is the code used for this tense as explained on p. 45.

What we have shown is the original structure of the data base. In order to facilitate searching, an extra attribute, EXAMPLE, was added later on, which combined the information from the SENTENCE and VERB attributes. In the case given the value of EXAMPLE would be 91.

The system allows one to manipulate the data base in various ways. For instance, by the command FIND a set of records can be selected which fulfil conditions formulated by the user, e.g. FIND LANGUAGE FR CATEGORY 21 (find all records from the French questionnaire with category label 21). By a TYPE command I can then get a list of all the attributes – or just a subset of my choice – of the records chosen. Other commands make it possible to sort the records in various ways and obtain statistics of different kinds. We shall look at the ways in which these possibilities were exploited in the following section.

The search for cross-linguistic categories

Having looked at the creation of the original data base, we shall now look at the analysis that I performed on that material in order to test

the hypothesis that there is a small set of cross-linguistic category types to which the overwhelming majority of language-specific categories can be assigned. Some words about the nature of this testing procedure are in order here. Given that I had at my disposal data about a restricted number of languages and that my data are – like any data – of limited reliability and validity, I cannot of course finally verify or falsify the hypothesis. What I set out to do can be characterized as playing a certain ‘game’ with the material, according to rules that I have chosen myself, but which are such that my readers will hopefully agree that if this game of patience comes out, as it were, it can be said that the material supports the hypothesis, as far as that is possible. The point of the game, then, will be to see to what extent one can group the language-specific categories of the material into sets the members of which have a sufficiently similar distribution for it to be plausible to say that they are manifestations of the same cross-linguistic category type.

Before going into the methods used for comparing categories, I shall discuss some delimitations of the original material that were necessary to make the undertaking practicable.

The concept of ‘major TMA category’

The coding of the TMA categories found in the questionnaires resulted in a rather large set of ‘labels’ – around one thousand, or approximately 15 per language in the sample. It was clear that it would be too ambitious a task to include them all in the cross-linguistic analysis, and this for several reasons.

In chapter 1, we discussed the problems connected with the description of TMA categories in ‘marked’ contexts, such as negated sentences, questions and passive constructions. Briefly, there are simply not enough examples in the questionnaire to make it possible to draw any safe conclusions in these areas. There is also the general problem of categories that have a low number of occurrences in the questionnaire: if a category occurs, say, once or twice in the whole questionnaire, it is not possible to say anything definite about its use: the element of chance is too great.

Another consideration is that when looking at the TMA system of a language, it easily becomes unobtrusive if all more or less peripheral categories are included in the study.

It thus became necessary to delimit the set of categories to be considered in the course of the investigation. To this end, I introduced the notion ‘major TMA category’: this notion is defined operationally, and indeed somewhat arbitrarily. To start with, I assumed that it was possible to sort out those categories that were not TMA categories proper – that is, things like case, person and number markers and so on, to the extent that they had been coded.

Admittedly, such a step is not unproblematic: it is clear that e.g. case marking (e.g. the object marking systems of Fenno-Ugric languages) and voice may involve TMA notions; yet, one would not like to draw in case and voice in general into the investigation. The problem here is that in order to be able to discard a category as being irrelevant to the TMA system, we would have to make a thorough investigation of its semantics; to have to do that for all categories is precisely what we are trying to avoid, however.

In addition, I formulated the following two conditions:

- (i) that the category should be manifested by at least six instances in the TMAQ.
- (ii) that it should be used in affirmative, declarative, non-embedded, active constructions.

It can be seen that (ii) automatically excludes voice categories (such as the passive) and things like negation and question markers. In addition, categories traditionally labelled moods will not in general be among the ‘major TMA categories’, since these, as pointed out in chapter 1, predominantly occur in embedded contexts. The major exception is QUOT, the category used to signal second-hand information (see chapter 5). It should be admitted that there are quite a few categories that are perhaps not moods in the traditional sense but which seem to have a ‘pragmatic function’ in that they express things such as the speaker’s attitude, whether the content of the proposition is new information etc. Examples are particles like *yo* in Japanese or *qad* and *laqad* in Classical and Modern Standard Arabic. (For a discussion of the latter, see Dahl & Talmoudi 1979.) In general, such categories were excluded from the final set of categories, a decision motivated more by the impossibility of drawing reasonably safe conclusions about universals in this area, due to the rather small number of categories of this type (and in addition, the fact that the questionnaire was not particularly well suited for finding out how they worked), than by any principled reluctance to regard them as TMA categories.

When starting the analysis, my working assumption was that the ‘labels’ used in the coding of the questionnaires could be identified with ‘TMA categories’. This is of course not necessarily the case, and in actual practice I sometimes had to re-analyse the material. For instance, I found that the strict ‘superficiality’ principle mentioned on p. 46 had to be modified somewhat: cases in point are the languages which have a lexically conditioned choice between several auxiliaries to mark the Perfect. In order to arrive at categories that were more suitable for cross-linguistic comparison, I decided to reduce them to one category for each language.

Cross-linguistic comparisons

In order to make cross-linguistic generalizations about categories, you obviously have to start by cross-linguistic comparisons. In material of the kind we are discussing here, a natural choice is to compare the distribution of categories between languages: the problem is just what categories should be compared and how.

One rather simple way is as follows: Given that you have a language-specific category C that interests you, take the set of contexts in which C occurs and investigate what categories are used in those contexts in all the languages of the sample. In doing so, look for language-specific categories that appear to have a similar distribution to the one you started with. Make a list of these and compute how many of them are used in each example in the questionnaire. Assuming that you have been able to find a sufficiently large number of language-specific categories the distributions of which are sufficiently similar, it makes sense to start talking about a cross-linguistically valid category type – what we have been referring to as a cross-linguistic TMA category.

However, what does 'sufficiently similar' mean? It is rather unsatisfactory to have to rely on intuitive judgements here; besides, it is very hard to compare by ocular inspection, as it were, two language-specific categories which may have up to a hundred occurrences in the questionnaire, and this becomes even more cumbersome if you have to make comparisons between, maybe, 50 different categories. It would be nice if the computer could do part of this job, and in addition provide an objective measure of similarity.

Such measures have been well known in statistics for a long time. The problem for the lay statistician is just to find the right one.

The so-called **contingency coefficient** 'C' (Siegel 1956, 196) is a measure of the extent of association or relation between two sets of attributes. It is basically defined as:

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$$

where N is the size of the total population under study and χ^2 is the standard chi-square measure. In our case, the total population would be the set of all verbs (250) in the questionnaire. (For further details, the reader is referred to Siegel or any similar textbook.) The upper limit of C is determined by the number of things we compare. In the cases that interest us primarily, we are comparing the distribution of two categories (call them A and B) in the questionnaires, and what we have is a 2×2 table, as in figure 2.3, that is, for each verb in the

questionnaire, there are four possibilities, each corresponding to one cell in the table: (i) both A and B occur, (ii) only A occurs, (iii) only B occurs, (iv) neither A nor B occurs.

+A +B	+A -B
-A +B	-A -B

Figure 2.3

In such a case, the upper limit of C, that is the value we obtain when the distributions are identical, is .707 (i.e., the square root of 2 divided by 2). (The cells (ii) and (iii) are then empty.) If there is no correlation, that is, in the perfect random case, the value is zero. It would probably be more convenient to have 1 as the maximum value, but this is hardly a serious problem. A clear disadvantage, however, is that C is not reliable when the expected value of the cells is less than 5, which happens fairly often in our investigation. As a way of correcting possible biases, I therefore decided to add a more simplistic but less vulnerable measure, the 'hit rate', which is described below.

Notice that what we measure when comparing two language-specific categories in this way is their distribution in our material: the coefficient does not give a way of comparing the actual distribution in the languages, if such a comparison makes sense at all (see the sampling discussion in chapter 1). We hope, of course, that the degree of similarity between two categories in our material in some way reflects how similar they 'really' are: to what extent this is true, we cannot tell; as usual, all the claims we make should be hedged: 'Judging from our material, the best possible guess is . . .'

As we said, making comparisons within a large set of language-specific categories is cumbersome; what we would like to be able to do is to make a comparison not between the members of every possible pair in the set, but rather between an individual category and the set as a whole. After all, what we want to arrive at in the end is a cross-linguistically valid characterization of a category and a list of language-specific categories that can be subsumed under it. In the ideal case, then, what we would like to be able to do is to compare each individual language-specific category to an 'ideal' distribution. However, how can we get at this 'ideal' distribution? Well, if we know which language-specific categories are subsumed under the cross-linguistic category, we could make a list of those examples that occur

most frequently with those categories. But this seems like a vicious circle: our original problem was how to identify the language-specific categories mentioned. Happily, the circle is only apparent. What has to be done is the following: starting out from a 'qualified guess' about what the ideal distribution will be, we look for the categories that conform best to that guess: having chosen that set, we are then able to make a better guess on the basis of how these languages are distributed; we can then see what languages conform best to that hypothesis etc. In the end, we will have the best possible hypothesis given the assumptions and the material at hand.

This was also how I proceeded. The first hypothesized cross-linguistic category that I tested this method on was the PERFECT: I wanted to see to what extent the much discussed Perfects of e.g. English and Swedish corresponded to something on the cross-linguistic level. I therefore chose a small set of languages (actually, English, Finnish and Swedish) which I knew had rather similar Perfects; I found the set of examples in which all these languages used a Perfect: I then went through all languages in the sample, finding which language-specific categories were used in these examples. (Needless to say, this was done on the computer.) This resulted in a list of hypothetical language-specific instances of PERFECT: I could then make a better list of examples, which could in its turn be tried on the categories and so on. (The first attempts were actually made before I introduced the 'C' similarity measure: I then just made a heuristic judgement of the similarities in each case.)

Later on, I found it expedient to take all categories with a traditional label, such as 'Progressive', to get the first guess at a hypothetical distribution. This approach may seem a bit shocking: some people suspect circularity here – am I not begging the question by assuming that these categories really represent the cross-linguistic category I am looking for? Again, this is only apparent. The point is that only if a reasonably large subset of the traditional Progressives have a common 'core' will it be possible to obtain a set of examples where these categories occur frequently enough to yield in the next step of the analysis significant values of the similarity measure. It is also reasonable to assume that there is some basis for the traditional label: names of categories are not given completely at random, even if this may sometimes seem to be the case.

Now a problem arises, however. It turned out that the number of occurrences in the questionnaire could vary quite considerably between the hypothesized language-specific manifestations of the same cross-linguistic category. For instance, among the hypothesized instances of the cross-linguistic category PROGRESSIVE, the frequency varies between 7 and 39. The question now is how to determine how large the 'ideal' distribution should be assumed to be.

It is clear that the degree of similarity between two sets will automatically have to be lower if they differ as to size.

This is where the prototype hypothesis comes in. The assumption that the cross-linguistic categories are defined primarily by their prototypical uses, and that there are variations among languages as to where the borderlines are drawn, suggests that if we find two language-specific categories which differ as to how often they occur in the material, this could be attributed to such variation. Furthermore, it suggests that we need not assume a fixed 'ideal' distribution for a category, if we assume that what is determined universally is, for each example in the questionnaire, the probability of occurrence, depending on the similarity between that example and the prototype of the category. This sounds a bit complicated, although it is not in fact. What it means in practice is the following. For each postulated cross-linguistic category C, we establish a ranking list of the examples in the questionnaire. The hypothesis is now simply that if C is manifested as a language-specific category C' and C' occurs n times in the questionnaire, those n occurrences will be exactly the top n examples on the ranking list. Thus, a PFCT category having 18 occurrences in the questionnaire will, in the ideal case, occur in the top 18 examples on the PFCT list. In reality, of course, this ideal case will very seldom be realized, but it constitutes the standard against which the similarity is measured.

Table 2.3 Distribution of *Kammu hóoc* construction

Number on cross-linguistic 'ranking list' of PFCT	Number of example in TMAQ	Is <i>hóoc</i> construction used?
1	531	yes
2	541	yes
3	561	yes
4	421	yes
5	641	yes
6	671	no
7	1361	yes
8	1392	yes
9	1341	yes
10	371	yes
11	1331	yes
12	1531	yes
13	1351	yes
14	1382	yes
15	431	yes

Table 2.3 continued

16	591	yes
17	481	yes
18	891	yes
19	681	no
20	511	no
21	1501	yes
22	571	no
23	521	no
24	1491	yes
<i>Limit of hypothesized distribution</i>		
25	901	no
26	551	no
27	1291	no
28	411	no
29	1072	yes
30	1082	no
31	691	no
32	491	yes
33	441	yes
34	631	no
35	1541	no
36	1301	no
37	1271	no
38	1281	no
39	611	no
40	1511	no
41	401	no
42	1051	no
43	581	no
44	621	no
45	601	no
46	501	no
47	1052	no
48	1061	no
49	1431	no
50	1481	yes
51	391	no
52	1062	no
53	1411	no
54	651	no
55	1132	no
56	381	no
57	1441	no
58	451	yes

I shall illustrate what I have said now by a concrete example. In table 2.3, I give a list of the 24 examples in which the Kammu 'hóoc' construction is found compared with the distribution that category would have in the above-mentioned ideal case, i.e. the top 24 examples on the cross-linguistic ranking list of PFCT. As we can see, there are 19 'hits' ('yes' above the dividing line) and 5 'misses' ('no' above the dividing line).

Since the sets to be compared now were by definition of equal size, it also became easier to compare them: it was here that I was able to introduce a simple 'hit rate' (HR) measure: it has the formula H/N , where H is the number of 'hits', i.e., the number of cases where the language-specific category was predicted to occur and actually did occur, and N is the total number of occurrences of that category. It turned out that the two similarity measures – C and HR – gave almost identical rankings of language-specific categories relative to the hypothesized distributions. HR also has its weaknesses – it is hard to compare the HR values for two pairs of categories (whether we are comparing two language-specific categories or one language-specific and one hypothesized cross-linguistic category) which are very different in size. I decided to keep both measures to enhance the reliability. For 'middle-sized' categories, the following rough correspondences between the measures obtain:

C	HR
.70	1.0
.60	.80
.50	.65
.40	.50

How do we arrive at the ranking list, then? The assumption here is that the prototypical uses of a cross-linguistic category correspond to those examples that are most often found with the language-specific manifestations of that category. This should not be seen as something that is assumed by definition or even that will necessarily always be the case, but it provides us with a useful heuristic for arriving at the ranking lists, viz. by counting language-specific categories. So, having found a set of such categories that we assume represent a common cross-linguistic category, we simply count, for each example in the questionnaire, the number of categories in the set chosen that are found with that example. The examples can then be ranked by the numbers we get. What we are trying to do is to establish, for each example in the questionnaire, its propensity to be used with a certain cross-linguistic category. Again, this is something that would be extremely laborious without a computer.

We can now check the list of language-specific categories to see what their correlations with the 'ranking list' are. In some cases, we may find that the correlation turned out to be too low: the

language-specific category should then be struck from the list. It may also turn out that some category that was not originally included had a high enough correlation to merit inclusion. The revised list of language-specific categories can now be used to give a new ranking list of examples, and so on, until we think that we have arrived at the optimal lists. Even at the computer, these calculations are relatively time-consuming (and therefore expensive!) and the number of rounds in this heuristic game must not be allowed to grow too much.

How similar to the hypothesized cross-linguistic category should a language-specific category be in order to be labelled as an instance of it? This is to some extent an arbitrary question: it depends on what amount of 'fudge' you allow for. Given the conditions under which the data were collected, there must be a significant amount of noise in the channel, as discussed above, which means that we should not expect to arrive at perfect correlations even in optimal cases. In addition, all systems allow for a certain free variation, which decreases the predictability of the output. In view of these factors, I think it is reasonable to regard C values above .60 or HR values above .80 as indications of optimal fit between the hypothesis and the data. Values below that, on the other hand, indicate that there is something going on that we do not understand. How seriously we should take that depends on several factors, however. If the values do not fall too much, we may still regard the hypothesis as reasonably good. I decided to draw the line at a C value of .40, which usually means that the hit rate is at least .50, that is, the hypothesis predicts at least half of the occurrences correctly. This may not sound too impressive, perhaps, but what should be taken into account is that the hypothesis we are talking about is a very strong one: that the occurrences of the category we are looking at should be exactly the top ones on the 'ranking list'. A category that occurs 10 times in the questionnaire and has a .50 hit rate may still have all its instances ranked among, say, the 20 top examples.

If a language specific category falls below the $C = .40$ level and thus does not qualify for identification with a cross-linguistic category CAT but its extension still falls within the 'upper half' of the ranking list of CAT, I give it the label \langle CAT, where the \langle sign is a typographically more convenient substitute for the inclusion sign (\subset).

We shall look closer at one category to see what this means in actual practice. In table 2.4, I have listed the set of categories that appear to fit the assumed ranking list for the cross-linguistic category PERFECT best. For each category, I have given its C value and the number of examples that fall into each interval of 10 on the ranking list.

We started from the working hypothesis that the number of languages that use a certain cross-linguistic category in a certain context shows us how close that context is to the focus of the category.

Table 2.4 Distribution of PFCT categories on the cross-linguistic ranking list

Language	Code of language-specific category	Value of C parameter	Number of occurrences in 10-interval below										Total frequency
			10	20	30	40	50	60	70	80	90	100	
Swedish	H	.64	10	10	10	10	7	6	6	1	1	0	61
Persian	Bu	.63	9	10	6	9	7	1	5	0	1	0	48
English	H	.61	10	9	8	4	4	2	0	0	0	0	37
Kammu	H	.61	9	8	3	2	1	1	0	0	0	0	24
Limouzi	H	.61	8	9	10	7	7	7	3	2	0	1	54
German	H	.59	9	9	8	9	6	7	4	3	5	2	64
Maori	pf	.58	10	6	6	1	2	1	0	0	0	0	27
Indonesian	S=	.57	8	5	5	0	1	1	0	0	0	0	21
Spanish	H	.56	9	7	6	7	8	6	1	4	1	2	55
Azerbaijani	22	.55	10	9	6	9	7	5	7	3	5	3	68
Işekiri	R	.54	6	7	4	0	0	0	0	0	0	0	18
Catalan	H	.53	9	7	6	8	7	5	2	4	2	2	56
Javanese	W	.53	8	6	5	2	0	2	0	0	0	0	24
Akan	p=	.52	7	6	4	0	1	0	0	0	0	1	22
Finnish	O	.51	8	8	7	7	8	5	5	0	4	1	62
Punjabi	A	.50	7	9	9	6	3	3	3	2	2	1	51
Sundanese	Pa	.49	9	4	4	0	2	2	1	0	0	0	23
Estonian	O	.49	9	7	7	6	4	4	1	1	4	0	49
Bulgarian	24	.49	9	4	5	7	3	1	0	2	0	3	35
Georgian	p	.49	7	10	5	5	3	1	7	1	0	2	45
Bugis													
Makassar	L	.49	8	5	6	5	1	2	1	0	0	0	28
Yoruba	p=	.47	7	4	6	2	1	3	0	0	0	0	26
Wolof	Na=	.44	8	4	5	0	1	0	0	1	0	1	29
Amharic	H1	.44	7	4	4	3	1	0	0	0	0	0	19
Tamil	I0	.41	6	5	4	2	5	2	0	0	0	0	24
Greek													
(Modern)	E	.39	4	6	5	3	3	0	0	0	0	0	21
Hindi/Urdu	A	.38	6	9	7	4	3	4	5	4	3	3	63
Kikuyu	it=	.37	4	3	5	2	1	0	0	0	0	0	16
Bengali	A	.35	10	9	4	5	4	6	2	6	6	5	80
Greenlandic													
Eskimo	2	.32	5	3	4	5	4	2	2	1	0	0	27
Oromo (Galla)	er	.22	2	0	5	0	1	0	0	0	0	0	8
Czech	M	.14	1	1	1	2	1	0	0	0	0	0	6
Portuguese	T	.12	2	3	1	1	2	1	0	2	0	0	12
Beja	C	.12	2	3	1	4	1	0	0	1	0	0	12

A somewhat different assumption is that the cross-linguistic focus or set of prototypical uses also corresponds to the focus of each language-specific category: that is, that it will also be true that the propensity for using each of these categories – computed separately – will be highest in the assumed focus area. In other words, the probability that the category in question is used will decrease monotonically as we go down the ranking list. As can be seen from table 2.4, this prediction is fulfilled quite well: at least in the upper half of the list, it appears safe to conclude that we have succeeded in identifying not only the cross-linguistic tendencies but also the foci of individual language-specific categories.

After having seen how well the language-specific categories that were identified as PFCT fit the cross-linguistic ranking list, the reader probably wants to know how well or badly they would fare if compared to other possible cross-linguistic categories. Maybe in fact that comparison would come out equally nicely? Let us therefore look at some cases to see if there is any risk of mis-identification.

'Perfect' and 'perfective' are two concepts that are not always kept apart, and many people may wonder if there is really any difference. In chapter 3, I shall discuss what I assume to be the cross-linguistic TMA category PERFECTIVE, and what the ranking list for this category looks like. We may now compare some of the PERFECT categories with the list for PERFECTIVE. The Swedish Perfect, for instance, turns out to have $C = .02$ and $HR = .23$ for PERFECTIVE. Kammu has $C = .06$ and $HR = .04$. To take an example further down the PFCT list, the Estonian Perfect has $C = .12$ and $HR = .10$. If we compare the lists themselves, it turns out that if we choose the top 28 examples on the PERFECT list, the hit rate when we take the top 28 examples from the PERFECTIVE list is .00! In other words, all obtained values seem very low, and the risk for confusion is rather small, at least in this case.

As we shall see in chapter 5, one category that is rather close to PERFECT is EXPERIENTIAL, whose meaning appears to be included in that of PERFECT. However, the ten top examples on the EXPERIENTIAL list get the values $C = .06$ and $HR = .10$, respectively. Even in this case confusion seems unlikely.

As we shall see later, however, there are cases of considerably higher correlations between cross-linguistic categories (see chapter 3 and the discussion of PROG vs. IPFV).

Some limitations of the 'ranking list' method

There are clear limitations to the method I have outlined here. One has to do with the size of the categories compared: it is clear that if the frequency of a category is too low, the element of chance becomes too great, but there are also difficulties when the category has too high a

frequency. Here is a concrete example: in Afrikaans, there is no PFCT category, but there is a general Past tense which formally looks like the Perfect in English. This category has a very high frequency in the questionnaire – 145 examples. If we calculate the C and HR values for this category and PFCT, they turn out to be .43 and .77, respectively. This seems to mean that the Afrikaans Past would, after all, qualify as a PFCT category! The problem is that when the frequency of the category is so high – over half the examples in the questionnaire – the hit rate almost inevitably becomes high, too, and when there is some relation between the categories we are comparing (after all, most typical PFCT examples are expressed by Past in Afrikaans), it may well happen that C also goes up – although C is less vulnerable to variations in category size than HR.

Another reason that large categories are problematic is that they commonly are 'default' categories in the sense that their application depends on the non-application of some other category or categories. Thus, in languages with instantiations of both PFCT and PAST, the category representing PAST is used if the conditions for using PFCT are not fulfilled. This makes it very hard to compare such categories cross-linguistically, since their distribution depends on what other categories there are in the language. For this reason, I have not tried to establish ranking lists for categories like PAST.

The conclusion is that the method I have outlined is best suited for 'middle-sized' cross-linguistic categories, i.e. categories with a normal frequency between 15 and 50 in the questionnaire.

It is also more suited for finding the foci of categories than for identifying variation in the occurrence of secondary uses. The reason is that each secondary use will be represented at best by a handful of examples in the questionnaire and if we do not have independent evidence about the use or non-use of a certain category in such contexts, the data usually do not suffice to draw any conclusions. Also, since examples representing secondary uses according to the hypothesis will appear rather far down the ranking list, the use or non-use of a category in such examples will have rather little influence on the C and HR values. We may exemplify this by the PERFECT category again: in chapter 5, we shall identify a variety that we shall call PFCT_q or PERFECT-QUOTATIVE, which differs from other instances of PFCT by being used also in 'quotative' contexts (see chapter 5 for further discussion). To this group belong Persian, Azerbaijani and Georgian. As can be seen from table 2.4, they do not have lower C and HR values than the others: Persian is actually second on the list.

A difficulty of perhaps greater theoretical significance is the fact that when talking about 'the propensity for a certain category to be used in a given context', what we are presupposing is a 'Boolean' model of how TMA categories are chosen. The best way of illustrating

what I mean by this perhaps slightly cryptic statement is to show what happens in the cases when the assumption is not true. Suppose for instance that a language has a Perfect and a Future. For any given context, we might then want to know how likely the use of these categories, or a combination of them, is. The 'Boolean' way of thinking deceives us, however, in that there are actually more than the 2×2 possibilities that there seem to be: a combination of the Perfect and the Future may be either a 'Perfect of the Future' or a 'Future of the Perfect', and the propensity to use these cannot be reduced to the product of the propensities of the individual categories. We are thus using a simplified model that may lead us astray if we think that it can be applied in all cases. This is important to bear in mind, but does not invalidate the method in general.

The relation between interlinguistic and intralinguistic variation

We have argued that the focus of a cross-linguistic TMA category will coincide with those examples in the questionnaire where the language-specific manifestations of that category are most often used. Data like those found in table 2.4 give support to this hypothesis, but we would be happier about it if we had more extensive data about individual languages. In general, it can be said that what the data we have collected provide us with is information about cross-linguistic variation, but since we have a very limited number of informants for every language, it is hard to say anything about variation within individual languages. Without starting a second investigation of the same size as the one I am reporting here, it is not possible to remedy this situation on a large scale, but some pilot probing is possible. In Lundberg 1984, some data are reported concerning the Swedish PFCT category (the auxiliary *ha* 'have' + the so-called supine of the main verb). Approximately 30 sentences from the TMAQ were translated into Swedish and presented to a group of Swedish speakers, asking them to choose between the Simple Past, the Present Perfect and the Pluperfect. The results are given in table 2.5. We can see that there is a strong correlation between the cross-linguistic rankings and the Swedish rankings, although in a number of cases the use of the Swedish category deviates from the hypothetical cross-linguistic distribution. In particular, we can see that there is an almost perfect agreement in the top group of the two lists. In other words, the hypothesized focus of the cross-linguistic category coincides with the focus of the language-specific category in this particular case. Lundberg's results are thus consistent with the hypothesis mentioned in the beginning of this section. We can also note another thing which is in accordance with our hypotheses, namely that there is a much steeper fall in the Swedish variation curve than in the cross-linguistic one. This means that there is less variation in the Swedish category

Table 2.5 Variation in the Swedish Perfect Category

1	2	3	4	5
53	31	19	yes	no
56	28	19	yes	yes
37	25	19	yes	no
133	24	19	yes	no
153	24	19	yes	no
135	24	19	yes	yes
43	23	6	yes	yes
48	22	16	yes	no
51	21	16	yes	no
52	21	18	yes	no
149	21	18	yes	no
129	19	9	yes	yes
41	19	18	yes	no
49	16	11	yes	yes
44	16	2	yes	yes
127	15	17	yes	yes
128	15	18	yes	yes
40	14	16	yes	no
60	13	7	yes	yes
50	13	2	no	yes
148	11	19	yes	no
39	10	9	yes	yes
65	9	16	yes	no
38	9	8	yes	yes
45	8	1	no	yes
66	7	17	yes	yes
47	7	2	no	yes

The numbered columns in this table indicate:

- (1) TMAQ sentence no.
- (2) No. of language-specific categories labelled PFCT which are used in the given sentence.
- (3) No. of informants in Lundberg's group who used a Perfect or a Pluperfect in the given sentence. (Total no. of informants: 19.)
- (4) Was a Perfect or a Pluperfect possible in this sentence according to the original informants? (Total no. of informants: 2.)
- (5) Was a Past tense possible in this sentence according to the original informants?

than in the cross-linguistic PFCT, something which is consistent with the claim that language-specific categories are sharpenings of the cross-linguistic ones.

Lundberg's investigation also gives us some indication of the reliability of the judgements in the questionnaires. We can see that, again, there is on the whole a fairly good correspondence between our original informants' intuitions and the judgements that Lundberg's

subjects gave. Needless to say, it would have been nice to have the same kind of check on some other language than Swedish, where the reliability of course can be expected to be higher than in other cases.

Are the ranking lists biased?

I have based the ranking lists on *all* the language-specific categories in the sample. The point of this arrangement is to obtain as large a number of instances as possible. It may give rise to areal and genetic biases, however, since one group of languages may be over-represented in the sample. Looking at the list of PFCT categories, some suspicion may arise from the fact that the top category on this list is Swedish, the native language of the author of the questionnaire. Indeed, if one looks at the C and HR values in general, they seem to tend to be higher for 'Standard Average European' languages. Some calculations give support to this feeling: if one computes the mean C value for all major TMA categories (that is, for the categories for which C has been calculated) of Indo-European languages spoken in Europe as compared to the corresponding values for all other languages, it turns out to be considerably higher: .60 as against .54. This may have several explanations: it may be that the questionnaire is indeed more suited for European languages in the sense that the examples fit better into the European categories, but it may also be that the amount of 'noise' is greater for the non-European languages in terms of proficiency in English among the informants and similar things discussed above. It should be kept in mind that even .54 is a fairly high value: it means that the hypothesis predicts considerably more than half the occurrences of the respective categories in the questionnaire.

In any case, it seems wise to test how large the bias in the ranking lists may be. To this end, I chose a sample of eleven from among the PFCT categories in such a way that as many language groups as possible were represented by at most one representative but excluding the Germanic group. The result was a slight reshuffling of the ranking lists of examples and categories: the C value of Swedish went down .03 (to .61) and it had to leave the top position to Kammu, a definitely 'exotic' language, whose C value went up to .63. Other changes were of the same order of magnitude. The C value obtained for correlation between the ranking lists themselves was at least .62 if the top ten examples were chosen and went up to .69 for the top 40 examples. In other words, there may well be a bias in the rankings, but it appears to be fairly marginal.

The problem of complex categories

Frequently, a sentence may contain several markers of TMA categories. For instance, the English sentence *He would have been swimming* contains at least the following TMA markers: the auxiliary *will*, the suffix *-d*, the auxiliary *have* (in combination with a past participle) and the copula *be* (in combination with an *-ing* form). This creates problems for the analyst: the main issue is whether the combination of two TMA markers should be regarded as the simultaneous manifestation of two TMA categories (one for each marker) or as the complex manifestation of one TMA category. Thus, the form *would* can be analysed as the past tense of the future auxiliary, i.e. as 'Future of the Past' or as an independent Conditional marker. This is a situation where you would like to have your cake and eat it, since it is often the case that the combined form exhibits both properties that are 'inherited' from the simple categories that make it up and 'emergent' properties, i.e. properties that are specific to it. The conditional in English and many other languages, for example, is used both in contexts where one would predict a 'Future of the Past' and in various special uses. In a prototype semantic framework, this should not come as a surprise: it was suggested above (p. 11) that secondary meanings often arise through the conventionalization of implicatures, and there is no reason why this could not happen also with complex categories. For the somewhat simplistic book-keeping system we have talked about in this chapter, the situation is troublesome, however, since there it appears that we have to make up our minds whether to count a form as a combination of two categories or as one complex category. If we choose the former alternative, we lose the possibility of being able to compare e.g. conditionals cross-linguistically; if on the other hand, we choose the latter alternative, we cannot e.g. compare future tenses in languages with and without past tenses. In actual fact, considerations of this kind have led to 'double accounting' in some cases: thus, pluperfects have been counted as instances of both the cross-linguistic category PERFECT and PLUPERFECT. Even if this is not wholly satisfactory, the conflict is there, and if one is aware of it, it should be possible to avoid any false conclusions.

Conclusion

This chapter has been a discussion of the methodological and practical problems connected with the investigation that lies behind this book,

and the principles that have been used in the search for cross-linguistic TMA categories. In the following chapters (3–5), we shall look at the actual outcome of this search, discussing each of the hypothesized cross-linguistic categories individually.

3

Aspectual categories

In this chapter, I shall discuss some categories which traditionally fall under the heading ‘aspect’. I shall argue that one can distinguish several types of such categories, the most important of which are PERFECTIVE:IMPERFECTIVE, PROGRESSIVE (PROG), and the group of categories treated under the heading ‘Habituals and generics’.

The opposition PERFECTIVE:IMPERFECTIVE (PFV:IPFV)

Perfectivity is often taken to be ‘the’ category of aspect: this is a view taken in particular by many Slavists, who are often unwilling to label anything that differs in any way from the Slavic opposition between Perfective and Imperfective as aspect. Also, the Slavic, or rather the Russian, aspectual system is often taken as a paradigm for what an aspectual system should look like. As I shall show in this chapter, if one looks at Slavic aspects from a typological perspective, it becomes clear that the Slavic systems are in fact rather idiosyncratic in many ways.

In various disguises, the PERFECTIVE:IMPERFECTIVE (PFV:IPFV) opposition occurs in about 45 of the languages in the sample. In table 3.1, a list of the categories that are probable candidates for being the marked members of the opposition in the different languages is given. In table 3.2, I list some parameters – to be discussed below – where there is cross-linguistic variation in the behaviour of PFV:IPFV. For comparison, the list also includes the accusative/genitive cases in Finnish and Estonian. Perfectivity plays an important role in the determination of direct object case marking in these languages: I do not want to claim that they are manifestations of the cross-linguistic category PFV.

PFV:IPFV differs from most other cross-linguistic TMA categories by not having very clear marking relations: in almost all other cases, we can choose one member of a TMA opposition as the universally marked one, and this choice will not be contradicted by more than one

Table 3.1 PFV/IPFV categories

Language	Code	Description	Frequency	C	HR
(i) PFV					
Maltese	p=	Perfective	106	.66	.92
Arabic (Tun.)	p=	Perfective (Past)	128	.65	.93
Italian	A	Passato prossimo	112	.64	.91
Latin	210	Perfectum	109	.64	.91
Bulgarian	21	Aorist	88	.63	.88
Portuguese	21	Simple Past	108	.61	.87
Greek (Modern)	p	Perfective aspect	104	.60	.86
French	21	<i>avoir (être) + V</i>	98	.59	.84
Limouzi	21	Past Simple	85	.58	.00
Alawa	22	Past Punctiliar	97	.57	.81
Georgian	22	Aorist	96	.57	.81
Beja	21	Preterite	94	.56	.80
Spanish	22	Simple Past	65	.56	.75
Hawaiian	p=	<i>ua + V</i>	132	.55	.84
Akan	2	Past	96	.53	.77
Seneca	Pi	Punctual indicative	113	.52	.79
Azerbaijani	21	Categorical Past (- <i>di</i>)	46	.51	.67
Arabic (Modern Standard)	21	Perfective	124	.50	.79
Turkish	210	Past Definite	70	.48	.67
Persian	2=	Past	79	.47	.68
Romanian	A	Aux. <i>avea</i> + Past Part.	88	.47	.69
Kurdish	21	Simple Past	88	.47	.69
Catalan	V	<i>va + Inf.</i>	56	.46	.63
Bandjalang	21	Past definite	84	.45	.67
Punjabi	21	PtP (predic.)	74	.44	.00
Chinese (Man.)	L	Particle <i>le</i>	64	.43	.61
(ii) PFV VOL					
Cebuano	rv	Prefix <i>mi-</i>	18	.10	.17
(iii) PFV N-VOL					
Cebuano	rpo	Prefix <i>naka-</i>	25	.15	.24
(iv) PFVd					
Bulgarian	p	Perfective verb	106	.52	.77
Russian	p=	Perfective verb	104	.51	.76
Polish	p=	Perfective verb	136	.49	.80
Czech	p	Perfective verb	130	.44	.75

Table 3.1 continued

(v) PFVd?					
Hungarian	p=	Perfectivizing prefixes	70	.11	.36
(vi) <PFV					
Romanian	21	Simple past	11	.32	.00
Oromo (Galla)	e=	Perfetto	85	.37	.60
Kikuyu	ir=	Suffix <i>-ir-</i>	13	.11	.15
(vii) IPFV					
Azerbaijani	11	Present	49		
Kikuyu	ag=	Suffix <i>-ag-</i>	11		
Oneida	s	Serial	37		
Turkish	110	Suffix <i>-yor</i>	48		
Cebuano	rdu	Prefix <i>ga-</i>	48		
Karaboro	h	High tone on subject	24		
Zulu	22!	Prefix <i>-be-</i>	99		

Table 3.2 Variation in PFV:IPFV

Language	1	2	3	4	5	6
Limouzi	-	-	-	-	-	-
French	-	-	-	-	-	-
Turkish	-	-	-	-	-	-
Akan	-	-	-	-	-	-
Kikuyu	-	-	-	-	-	-
Latin	-	-	-	-	-	-
Maltese	-	-	-	-	-	-
Bulgarian (Aorist)	-	-	-	-	-	-
Karaboro	-	-	-	-	-	-
Portuguese	-	-	-	-	-	-
Hawaiian	-	-	-	-	-	-
Azerbaijani	-	-	-	-	-	+
Italian	-	-	-	-	-	+
Amharic	-	-	-	-	+	-
Arabic (Tunisian)	-	-	-	-	+	-
Persian	-	-	-	-	+	-
Wolof	-	-	-	-	+	+
Oneida	-	-	-	-	+	+
Beja	-	-	+	-	-	-
Arabic (Mod.Stand.)	-	-	+	-	-	-

Table 3.2 continued

Oromo (Galla)	-	-	+	-	-	+
Kurdish	-	-	+	-	+	-
Punjabi	-	-	+	-	+	-
Zulu	-	-	+	+	-	-
Sotho	-	-	+	+	-	+
Romanian (Rumanian)	-	+	-	-	-	-
Seneca	-	+	-	-	-	-
Alawa	-	+	-	-	-	-
Spanish	-	+	-	-	-	-
Catalan	-	+	+	-	-	-
Georgian	-	+	+	-	+	+
Greek (Modern)	-	+	+	+	-	-
Chinese (Mandarin)	+	-	-	-	-	+
Hindi/Urdu	+	+	-	-	-	+
Bandjalang	+	+	-	-	-	+
Bulgarian (Perf.)	+	+	+	-	-	-
Cebuano	+	+	+	-	-	+
Czech	+	+	+	+	-	+
Polish	+	+	+	+	-	+
Russian	+	+	+	+	-	+
Estonian	+	+	+	+	+	-
Hungarian	+	+	+	+	+	-
Japanese	+	+	+	+	+	+
Finnish	+	+	+	+	+	+

The column numbers show:

- (1) IPFV only used in (Q.13).
- (2) IPFV only used in (Q.95).
- (3) PFV used in subordinate clauses with future time reference.
- (4) PFV:IPFV opposition in main clauses with future time reference.
- (5) PFV used for 'reportive present'.
- (6) IPFV only used in (Q.29).

or two counterexamples. In the case of PFV:IPFV, on the other hand, it seems rather to be a typical situation that even in individual languages, we cannot choose one member of the opposition as being clearly unmarked. This is the reason why I have chosen to treat PFV:IPFV as an 'equipollent' opposition in the sense of European structuralism, although this term could be misinterpreted to imply that the two members of the opposition are equivalent, which is certainly too strong a statement.

The view taken here differs from that formulated in Friedrich 1974, where what we have labelled here PFV:IPFV is subsumed under a 'durative/non-durative' opposition, where the durative member is said always to be the more marked one. I think that the picture is skewed in Friedrich's treatment by his identification of PFV:IPFV and

Table 3.3 Distribution of PFV

Rank no.	No. of categories	Examples
1	32	911 1651 1751
4	31	921 991 1012 1621
8	30	141 1002 1671 1721
13	29	451 461 931 941 1271 1372 1631 1642 1741
21	28	871 1701 1731 1901
25	27	951 1411 1421 1641 1801
30	26	131 551 1301 1441 1681 1691 1692
37	25	381 391 491 881 961 1121 1351 1381 1491 1742 1771 1881
49	24	471 1131 1281 1321 1331 1371 1431 1831 1851 1871
59	23	291 561 1001 1141 1211 1391 1501 1531 1561 1781 1791 1792 1821 1842
73	22	541 601 641 681 691 882 1011 1401 1841 1892
83	21	571 621 1341 1361 1891
88	20	661 671 1392 1541
92	19	611 891
94	18	441 531 591 651 872 1291 1511
101	17	431 631 901 1471
105	16	371 1051 1382
108	15	282 421 501
111	14	411 481 521
114	13	1052 1061 1941
117	12	511 981 1191 1322 1931
122	11	281 1181
124	10	321 331 351 701 1021 1031 1041 1072 1161 1182 1961 1970

PROGRESSIVE: for PROGRESSIVE, the marking relations are much clearer in that the 'durative' member of the opposition is in fact always marked.

The difficulty of deciding which member of the opposition is marked and which is unmarked is connected with the tendency for PFV:IPFV to be realized not by affixation or by periphrastic constructions but rather by less straightforward morphological processes. The 'classical' Indo-European and Semitic systems are illustrations of this.

Semantics of PFV:IPFV

In the literature, there have been many attempts to characterize semantically what underlies perfective:imperfective oppositions.

Most of these attempts have aimed at finding a common feature – a ‘Gesamtbedeutung’ – of the perfective members of the oppositions – something which has appeared natural in view of the more apparent ‘polysemous’ character of imperfective categories. If the ‘prototypical’ approach to the semantics of grammatical categories that I am advocating is correct, the search for ‘Gesamtbedeutungen’ will in general be a futile one. The difficulty aspectologists have had in coming to an agreement on the question, What is the essential characteristic of perfectivity?, is understandable if they simply have been looking at different features that make up the description of the prototypical uses of PFV. It may also well be that although the prototypical uses are common to all PFV categories, there is variation among languages in how they delimit PFV. In fact, our material gives some support to such a hypothesis. Let us look at some critical cases, taking as a point of departure one rather common view of the nature of perfectivity as summed up by Comrie (1976, 16):

perfectivity indicates the view of a situation as a single whole, without distinction of the various separate phases that make up that situation, while the imperfective pays essential attention to the internal structure of the situation.

Let us call this the ‘totality’ view of perfectivity. In Dahl 1980a, I pointed out that this view is not equally adequate for all language-specific categories that it has been applied to. Consider (Q.9) and (Q.13–14):

(Q.9)

Context: A: I went to see my brother yesterday. B: What he DO?
(=What activity was he engaged in?)

Sentence: He WRITE letters

(Q.13)

Context: A: What did your brother do after dinner yesterday?

Sentence: He WRITE letters

(Q.14)

Context: A: What did your brother do after dinner yesterday?

Sentence: He WRITE a letter

The crucial sentence here is (Q.13). At least *prima facie*, it looks like a clear case of perfectivity according to the characterization above, and indeed in most languages with a PFV:IPFV opposition it is translated using a PFV form, as can be seen from table 3.2, although there are quite a few exceptions – a notable group being the Slavic languages (except for the Aorist in Bulgarian!). Let us look closer at

what happens in Russian. The translations of the sentences above are the following:

(Q.9:RU) On pisał pis'ma

(Q.13:RU) On pisał pis'ma

(Q.14:RU) On napisal pis'mo

The perfective counterpart to (Q.13:RU) would be:

(3.1) On napisal pis'ma

which is unacceptable if *pis'ma* is not interpreted as referring to a definite set of letters. The reason that a Perfective verb cannot be used in (Q.13) is that ‘writing’ is an unbounded activity as long as we have not delimited the object in any way. If we do so, however, as in (Q.14), the Perfective *napisal* can be used. In other words, the notion of boundedness is crucial to the choice between Perfective and Imperfective in Russian. To take another example, consider the following:

(3.2) Včera posle obeda my tancevali/potancevali

‘(What did you do yesterday after dinner?) We danced/danced a little’

Here, it is possible to use both an Imperfective verb (*tancevali*) or a Perfective one (*potancevali*), but the latter delimits the activity quantitatively (‘we danced a little’). Again, the Perfective aspect entails a bounded activity.

In Slavic aspectology, one often talks of the so-called ‘general factual’ or ‘simple denotative’ use of the Imperfective aspect, where ‘there is no specific reference to the completeness of the event’ and ‘the speaker is simply interested in expressing the bare fact that such and such an event did take place, without any further implications’ (Comrie 1976, 113). (Q.13) might be an example of this; another would be (Q.53), which also is translated into Russian with the Imperfective:

(Q.53)

Context: A: I want to give your brother a book to read, but I don't know which. Is there any of these books that he READ already? B:

Sentence: (Yes,) he READ this book

(Q.53:RU) On čital ètu knigu

Comrie says that the existence of such a use of the Imperfective is ‘perhaps the strongest single piece of evidence in Russian... for considering the Perfective to be the marked form’, the argument being that the Imperfective has a less specific interpretation since it

'may or may not have imperfective meaning'. The question is if such an explanation of the 'general factual' use of the Imperfective is compatible with the totality view of perfectivity – it would otherwise seem that the description given of this use makes it look very much like a prototypical case of viewing an event as a single whole without paying 'essential attention to the internal structure of the situation'. Referring to the absence of 'further implications' in the Imperfective (apparently as opposed to the presence of such implications in the Perfective) is in fact coming fairly close to an acknowledgement that the crucial factor is not 'totality' but something else.

The view that boundedness – or rather the attainment of a limit – is the crucial factor for aspect choice in Russian has been taken before (see e.g. Vinogradov 1947). Some authors make no difference between 'totality' and 'boundedness', however, and it is therefore worthwhile pointing out that if we make such an identification we cannot invoke the concept of totality to explain the difference between (Q.9) and (Q.13) and why in fact most languages draw the line between IPFV and PFV here rather than between (Q.13) and (Q.14).

In fact, Comrie's characterization of the difference between 'perfective' and 'imperfective' is not wholly adequate even for those languages where 'totality' comes closer to describing what is going on in PFV:IPFV. 'Paying attention to the internal structure of the situation' is a rather cryptic formulation which may be understood in various ways. Looking at a typical case of imperfectivity, say (3.3), it is however not at all clear in what way it could be said that it pays such attention.

(3.3) John was sitting in a chair

It is correct that in (3.1) we are asked to consider a part of the sitting process rather than the whole of it, but it may legitimately be asked if the whole process – and a fortiori its internal structure – is relevant at all.

The parallels between PFV:IPFV and the distinction between count and mass nouns has often been pointed out: it appears natural to say that a count noun denotes an object 'as a single whole', but it is a much less plausible statement to say that a mass noun 'pays essential attention to the internal structure' of an object.

A much discussed problem is that of the co-occurrence restrictions between aspects and adverbial phrases specifying length of time. Adverbials of the type *for two hours* are often thought of as most readily co-occurring with imperfective verb forms. This is certainly true of Russian, where a sentence like (3.4) is downright ungrammatical with the Perfective aspect.

(3.4) On pisał/*napisal knigu dva mesjaca 'He wrote the book for two months'

Both in Russian and other languages, however, counterexamples to this can be found, and sentences like the following belong to the stock examples of aspectological literature:

(3.5) On postojal tam čas 'He stood (Perfective) there for an hour' (Russian)

(3.6) Ebasileue deka etē 'He reigned (Aorist) for thirty years' (Classical Greek)

There appears to be a difference here, though, in that the addition of the perfectivizing prefix *po-* to the imperfective Russian verb *stojat*' (and many others) results in an additional nuance of 'shortness of duration' which is not present in analogous examples from many other languages. (Other prefixes may change the interpretation in other ways, e.g. the so-called 'perdurative' prefix *pro-* whose use rather seems to imply that the time interval referred to was long or in some way remarkable – the generalization that the perfective constructions are a marked choice remains valid, however.) In our questionnaire data, (Q.95) would be a relevant example:

(Q.95)

Context: Q: What your brother's reaction BE when you gave him the medicine (yesterday)?

Sentence: He COUGH for an hour

Among the categories listed in table 3.2, about half seem to choose the IPFV alternative here: in the Slavic group, only the Bulgarian Aorist:Imperfect distinction belongs to this group.

As a further example of a context where the Slavic languages use the imperfective aspect although many other languages have PFV, consider (Q.29):

(Q.29)

Context: Q: Did your brother finish the letter quickly? A:

Sentence: (No,) he WRITE the letter slowly

The Russian translation here is:

(Q.29:RU) On pisał pis'mo medlenno

The use of a perfective verb is hardly possible at all here. In general, many manner adverbials tend to co-occur with the imperfective aspect in Russian and the other Slavic languages. Traditionally, such cases are said to fall under the generalization that the imperfective aspect is used when the action itself – rather than its result – is relevant. (Notice that *On napisal pis'mo bystro*, 'He wrote the letter fast', is quite all right with the perfective aspect, meaning that the result was attained quickly.) The majority of the non-Slavic languages in the material do not show any reluctance to use PFV forms here. Not unexpectedly, Bulgarian uses the Imperfective Aorist.

H
PFV

Table 3.4 Prototypical occurrences of PFV

Sent. 175 No. of categories: 32 Do you know what happened to my brother yesterday? I saw it . . . (narrative) . . . It DIE
Sent. 165 No. of categories: 32 Do you know what happened to me yesterday? . . . (narrative) . . . It DIE
Sent. 91 Verb 1 No. of categories: 32 Q: What your brother's reaction BE to the medicine (yesterday)? He COUGH once
Sent. 162 Verb 1 No. of categories: 31 Do you know what happened to me yesterday? . . . (narrative) . . . Suddenly I STEP on a snake
Sent. 99 Verb 1 No. of categories: 31 Q: How long did it take for your brother to finish the letter? He WRITE the letter in an hour
Sent. 101 Verb 2 No. of categories: 31 Last year, the boy's father sent him a sum of money When the boy GET the money, he BUY a present for the girl
Sent. 92 Verb 1 No. of categories: 31 Q: What your brother's reaction BE to the medicine (yesterday)? He COUGH twice
Sent. 100 Verb 2 No. of categories: 30 The boy's father sent him a sum of money some days ago and it arrived yesterday When the boy GET the money, he BUY a present for the girl

Returning now to the question how the prototypical PFV can be characterized, we could give the following tentative description with support from the examples in table 3.4, in which the sentences where the assumed PFV categories are most often used are listed:

- (3.7) A PFV verb will typically denote a single event, seen as an unanalysed whole, with a well-defined result or end-state, located in the past. More often than not, the event will be punctual, or at least, it will be seen as a single transition from one state to its opposite, the duration of which can be disregarded.

A special problem concerns the realization of PFV:IPFV in habitual contexts. Mønnesland (1984, 54), questioning the traditional view in

Slavic aspectology that 'iterative is one of the two main functions of the imperfective aspect', points out that Slavic languages differ in how they solve the dilemma that arises when the 'single event:repeated event' criterion contradicts the 'total:non-total' criterion. In speaking of a repeated total event 'one can use a pf. verb, thus stressing each individual total event, or use an ipf. verb, which means that the stativeness of unlimited repetition takes precedence'. The first solution is normally chosen in Russian, Polish, and Bulgarian, whereas the second is preferred (even if it is not always the only possible alternative) in Czech, Slovak, Sorbian (a West Slavic language spoken on the territory of the German Democratic Republic) and Slovene. In Serbo-Croatian, both aspects are possible. In our material the only clear example of an 'iterative' Perfective is the following Czech sentence:

- (Q.102:CZ) Kdý chlapec dostane peníze, koupí vždy dárek pro dívku
'When the boy got (pf. pres.) the money, he always (word added in the translation) bought (pf. pres.) a present for the girl'

An example from Slovene from Mønnesland's paper is given here in shortened form:

- (3.8) Vsako jutro sem pozno vstal 'Every morning I got up late'

Even in Russian, where the Imperfective is normally used in habitual contexts, the addition of a nuance of 'potentiality' makes the Perfective possible in many cases:

- (3.9) On ne ošibetsja 'He won't be mistaken'

In terms of the semantic framework presented in chapter 1, it seems clear that habituality is a factor that enters into the semantics of PFV:IPFV but which may be given different weight in different languages. Such an account appears preferable to one where habituality is just one of several kinds of imperfectivity.

Restrictions on time reference in PFV

There is a strong tendency for PFV categories to be restricted to past time reference. I interpret this restriction as a secondary feature of PFV in the sense of chapter 1. In other words, for all languages it holds that 'past time reference' characterizes prototypical uses of PFV – single, completed events will in the 'typical cases' be located in the past. Languages will differ, however, in the extent to which they allow uses of PFV with non-past time reference. Also, within one and the same language, the 'past time reference' restriction may hold with unequal force in different contexts. A good illustration of this is the

Aorist in Homeric Greek, as described in Friedrich 1974. After noting that indicative aorists 'usually seem to refer to the past' he goes on:

The aorist also lacks obligatory past reference in the oblique moods and certain other, less important cases. The aorist participle, although it usually refers to an antecedent act, is obligatorily marked only for aspect. We may thus agree with Wackernagel that, 'Im ganzen ist zu leugnen, dass dem Aoriststamm an und für sich ein (sic) Bezeichnung auf das Präteritum anhafte.' In other words, past tense is an implication of the aorist forms, but it ranges from a limited probability to a weak connotation to zero. The one exception is the fact that the aorist necessarily refers to the past in contrary-to-fact constructions.

Likewise, in Modern Standard Arabic, the category variously called 'Perfect' or 'Perfective' normally has only past time reference but may refer to the future in certain subordinate clauses, as in the translation of (Q.104):

(3.10) ʔida: ʔit:asala elwaladu bilwa:li sayaštari:
 if comes the boy to the money he will buy
 hadiy:atan lilbin
 a present for the girl

'If the boy gets (Perfective) the money, he will buy (Future) a present for the girl'

The same seems to be possible in some other languages (see table 3.2).

There are, however, several exceptions to the tendency to restrict PFV to past time reference, the most notable being the Slavic Perfective. (Other examples are Japanese, Modern Greek and some Bantu languages, e.g. Zulu and Sotho.) Interestingly, though, the category of Perfective in e.g. Russian is not without its restrictions as to time reference: it is well known that the Russian Perfective non-Past normally cannot be used to refer to the moment of speech. A sentence like (3.11) will therefore be understood as referring to future time in its most natural reading:

(3.11) Ja napišu pis'mo 'I write (Perfective Non-Past) a letter'

Thus, Russian and the languages discussed above have in common that PFV is not possible with present time reference but differ as to the status of the future. There might thus be a strong and a weak version of the time reference restriction on PFV: it may be interpreted either as 'only past' or as 'only non-present'.

The fact that the Russian Perfective cannot have present time reference is sometimes interpreted as a necessary consequence of the semantics of this aspect. It is said that a Perfective verb form must refer to a 'total' event and that what happens at the moment of speech

is necessarily on-going rather than completed. A Perfective Present would therefore be a contradiction in terms. However, there are candidates for 'total' events at the moment of speech: notably the so-called reportive present and performatives. In fact, PFV is used in such cases in several languages, as can be seen from table 3.2. This suggests that these languages have a very loose time reference restriction on PFV, if any at all.

Perfectivity and tense

So far, we have spoken about the relation between perfectivity and time reference. We shall now consider the relation between PFV:IPFV and tense. It is of some importance to keep apart two things:

- (i) possible restrictions on the time reference of particular aspectual categories;
- (ii) the ways in which aspectual and temporal categories may be combined to build up a TMA system.

This distinction is not always made. Consider the following passage (Comrie 1976, 71):

One of the most interesting relationships between aspect and tense . . . occurs when an aspectual distinction is restricted to one or more tenses . . . It appears that the tense that most often evinces aspectual distinctions is the past tense. Thus in many Indo-European languages, and also in Georgian, the difference between the Aorist and the Imperfect exists only in the Past Tense, and there is no corresponding distinction in other tenses . . .

Comrie is here referring to the fact that it is very common for non-future TMA categories to form a tripartite system, as in the 'classical' Indo-European TMA system (neglecting the Perfect for the time being) shown in figure 3.1.

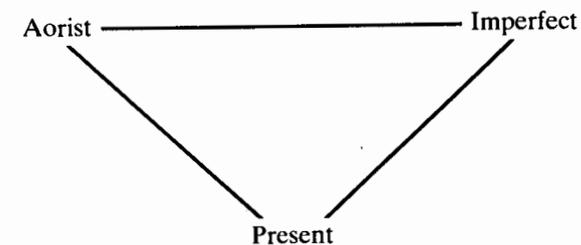


Figure 3.1

Comrie's formulations suggest an analysis of these systems as in figure 3.2.

	Pfv. aspect	Ipfv. aspect
Past tense	Aorist	Imperfect
Present tense	Present	

Figure 3.2

However, there are several facts that suggest that such an analysis is wrong. From figure 3.2 we would expect there to be a marker of past tense common to the Aorist and the Imperfect and some aspectual marker distinguishing them. In fact, this situation is not very common: it is questionable if it is found at all, at least not in the form described. If we look at the Indo-European languages, we see that instead, the Imperfect and the Present typically go together in that they are formed from the same stem (commonly called the Present stem) and that they are distinguished by different sets of personal endings. This suggests an alternative analysis (figure 3.3).

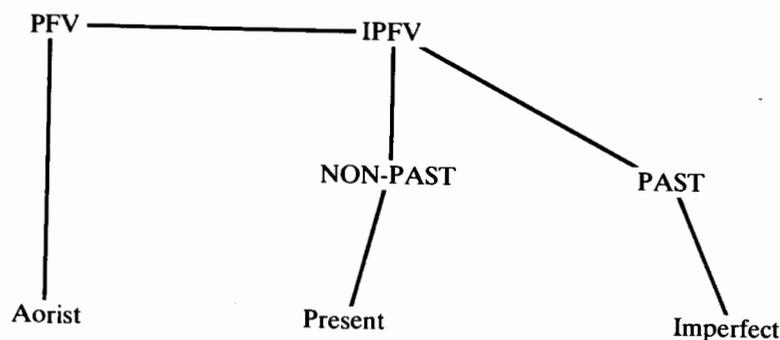


Figure 3.3

In such a system, there would be no 'Past Tense' category comprising both Aorist and Imperfect: Aorist is simply PFV, and the fact that it (normally) has past time reference is in accordance with the general tendencies for that category. Imperfect, on the other hand, would be analysed as a combination of IPFV and PAST. In addition to morphological arguments, such an analysis would be supported by the fact that (as pointed out in the quotation from Friedrich above) the Aorist is not always restricted to past time reference in all contexts. Furthermore, the analysis proposed here makes the classical Indo-European system look very much less different from the 'Semitic'

one, in which there are also three categories: (i) the 'Perfect' (or 'Perfective') characterized by suffixed subject markers (ii), the 'Imperfect' characterized by prefixed subject markers, and (iii) the imperfective past construction formed by combining a past tense copula with an Imperfect verb, as illustrated by the following example from Classical Arabic:

Perfective: *kataba* 'he wrote'

Imperfective: *yaktubu* 'he is writing'

Imperfective Past: *ka:na yaktubu* 'he was writing'

The systems would differ mainly in how PAST is marked and in to what extent the use of PAST is obligatory.

We notice, however, that figure 3.3 is not adequate for Classical Greek: contrary to what we have said about the general tendency in Indo-European, there is a consistent marker of past time reference in that language, the so-called augment, i.e. a prefix *e-* which characterizes both the Aorist and the Imperfect. How does this fit into the general scheme? This is the point where the distinction made in the beginning of the section becomes crucial. In this particular case, we must keep apart two things:

- (i) whether a PFV category is restricted to past time reference;
- (ii) whether the use of a PAST is restricted to IPFV.

Since these two are logically distinct, there are theoretically four possibilities, all of which may very well turn out to be realized in actual languages. The most common case is no doubt the one where (i) and (ii) go together in that both are answered by yes or both by no, but cases where they differ may also be found, and Classical Greek seems to be a case in point: the PFV category – the Aorist – is (in the indicative at least) restricted to past time reference, but the manifestation of PAST applies to both PFV and IPFV. The converse situation is less easily documented, at least if we are looking for a language where there is no time reference restriction at all on PFV – we do not have any unequivocal examples of those in the material. The languages mentioned above where PFV can be used for reportive presents are possible candidates, however.

In spite of the existence of cases like Classical Greek, the tendency for (i) and (ii) to go together is undeniable. There is an obvious functional explanation for it: if the mere use of a PFV form entails past time reference, further marking of past time reference is redundant. Interestingly, there is an exactly parallel situation with regard to future time reference in some of the languages which have the weaker variety of the time reference restriction on PFV. In a language like Russian, it is only the Imperfective aspect that has a special Future tense, viz. a periphrastic construction where a Future copula (*bud-*

with the appropriate ending) is combined with the infinitive of an Imperfective verb.

The nature of Slavic-style aspect

In the preceding section, we noted significant differences in the semantics of the Russian Perfective/Imperfective opposition and PFV:IPFV in a number of other languages. Interestingly, Slavic aspect has already been singled out a couple of times in this chapter as being different from the 'normal' cases of PFV:IPFV. Let us therefore have a closer look at these systems.

Let us review the basic facts of the Russian aspect system for the benefit of those who are not familiar with Slavic languages. Most Russian verbs can be classified as either Perfective or Imperfective. Morphologically nonderived verb stems tend to be Imperfective; there is, however, a sizeable number of exceptions, such as *dat* 'give', *leč* 'lie down', which are Perfective. (These verbs in general denote punctual or bounded actions.) There is in Russian a rather large set of verbal prefixes: these in general correspond semantically to the 'verbal particles' of Germanic languages, such as *up*, *down*, *out* etc. in English. Adding such a prefix to an Imperfective simple verb normally makes it perfective; thus, from the Imperfective *lit* 'pour', one can form the derived Perfective *vylit* 'pour out'. The prefixes originally had local or similar meanings; most of them can still have those, in addition to various secondary uses. Some prefixes, however, can also be used as 'empty' perfectivizers, that is, they do not seem to have any effect on the meaning of the verb except for making it Perfective. A hot issue in Russian aspectology has always been the nature of those 'empty prefixes', some linguists claiming that there is always some additional change of meaning in addition to the aspect change.

A Perfective verb, whether derived or basic, may be imperfectivized by adding to the stem a derivational suffix, most commonly *-va-* or *-yva-*. Historically, these suffixes seem to have had what is usually called an 'iterative' meaning; in our terminology, they most probably represented the category HAB.

Both Perfective and Imperfective verbs can have Past and non-Past forms; they differ, however, with respect to the periphrastic Future construction, which can only be used with Imperfective verbs. As a compensation, Perfective non-Pasts normally have future time reference.

We thus obtain paradigms of the kinds shown in table 3.5. Originally, of course, the Slavic languages had TMA systems of the classical Indo-European type with PFV:IPFV manifested as an opposition between Aorist and the other categories. What appears to have happened then was a grammaticalization of certain verb

Table 3.5

	<i>Basic imperfective verb</i>	<i>Prefixed perfective verb</i>	<i>Derived imperfective verb</i>
non-Past	l'ju 'I pour'	vyl'ju 'I will pour'	vylivaju 'I pour out'
Past	lil 'I poured'	vylil 'I poured out'	vylival 'I poured out'
Future (periphr.)	budu lit' 'I will pour'	—	budu vylivat' 'I will pour out'

prefixes, together with a reinterpretation of the originally 'iterative' suffixes. In most Slavic languages the Indo-European system disappeared via the expansion of an old PFCT construction combined with the complete disappearance of the old Aorist and Imperfect forms. In our material, this development is represented by Russian, Polish and Czech. In Bulgarian, on the other hand, the old Indo-European system was preserved alongside the younger, specifically Slavic system. The two Bulgarian systems are in principle independent of each other, although there is a rather strong tendency for the Aorist to be combined with the Perfective aspect and for the Imperfect to be used with Imperfective verbs.

As we saw earlier, the Slavic opposition is much more independent of tense and time reference than the corresponding categories in other languages; it also has a much more 'derivational' character than the average aspectual opposition. The question that arises is whether the Perfectivity/Imperfectivity opposition in Russian, Polish, Czech and Bulgarian should be subsumed under PFV:IPFV at all. In particular for Bulgarian, where Perfective/Imperfective exists alongside the Aorist/Non-Aorist opposition, this would appear a natural conclusion. However, there are also arguments that speak against at least a total separation of Slavic aspect and PFV:IPFV. If we look at table 3.2, it becomes clear that the Slavic categories differ from the majority on several parameters, but that on each separate parameter there are also other languages that behave like the Slavic. Furthermore, the correlations between the Slavic categories and the hypothesized distribution of PFV are still fairly high; in particular, if we compare the past tenses of the Slavic Perfectives, we obtain very high correlations indeed, as can be seen from table 3.6.

To get a better understanding of Slavic aspect, it is of some interest to look at other, similar systems in non-Slavic languages.

Morphemes with a function similar to those of the Slavic verbal prefixes occur in many languages, although they are not always realized morphologically in the same way: in English, for instance,

Table 3.6

Language	Correlation of Perfective with PFV	Correlation of Past Perfective with PFV
Russian	.51	.59
Polish	.49	.63
Czech	.45	.59
Bulgarian	.52	.60

they are, as we have seen, most often free particles. The semantics of such morphemes gives them a natural connection to aspectual categories: to take a typical example, *eat up* will be less naturally used in imperfective contexts than the bare *eat*. Still, it does not appear to be so common for this tendency to lead to the kind of grammaticalization of these morphemes that we encounter in the Slavic languages. In the literature, such tendencies are commonly reported outside of Slavic from the Baltic group (Latvian, Lithuanian), Fenno-Ugric (mainly Hungarian) and Georgian. Of these, Hungarian and Georgian are represented in our material. Let us first look at Hungarian.

In Hungarian, there is a set of aspect marking morphemes which resemble the Slavic perfectivizing prefixes but which may be either free (particles) or bound (prefixes). For instance, the prefix *el-* distinguishes the translations of (Q.53-4):

(Q.54)

Context: A: It seems that your brother never finishes books.

Sentence: (That is not quite true.) He READ this book (=all of it)

(Q.53:HU) Ezt a könyvet olvasta

(Q.54:HU) Ezt a könyvet elolvasta

These morphemes have been referred to as perfectivizing, and indeed they seem to be used mainly in contexts where PFV could be expected; although perfectivity is relevant, it appears that we are not dealing with a systematic manifestation of PFV: the correlation with the hypothesized ideal distribution of PFV is almost nil. One reason seems to be that their distribution is highly lexically determined: many verbs do not seem to have any aspectually marked counterparts. Also, it is clear that factors usually not relevant or at least peripheral to aspectual choice enter into the picture: for instance, definiteness appears to be crucial in many contexts. Consider the following translation of (Q.103):

(Q.103)

Context: The boy is expecting a sum of money

Sentence: When the boy GET the money, he BUY a present for the girl

(Q.103:HU) Ha a fiú megkapja a pénzt, ajándékot vesz a lánynak.

Here, an aspectually marked verb is used only in the temporal subclause, but if we change the definiteness of the objects in the respective clauses, the aspectual marking will also change, so that we would get the following variants:

(3.12) Ha a fiú pénzt kap . . . 'If the boy gets money . . .'

(3.13) . . . megveszi az ajándékot a lánynak ' . . . he will buy the present for the girl'

A non-default combination of aspect and definiteness demands a special context and will have a special interpretation. For instance, *megvesz egy ajándékot* 'buys a present' (perfective + indefinite) will focus on the action of shopping rather on getting a present for someone. *Veszi az ajándékot* could be used in a temporal clause in the sense 'when he is buying a present'. Another use would be in a scornful statement like the following:

(3.14) Nincs pénze, de veszi az ajándékot. 'He hasn't got a penny, but he buys presents'¹.

These observations certainly do not exhaust the problem of the function of the Hungarian perfectivizing morphemes: for further discussion and references, see de Groot 1984.

Another language with a 'Slavic-like' aspect system in our material is Georgian, which also has a set of perfectivizing prefixes. Like Bulgarian, Georgian has an opposition between Aorist and non-Aorist forms; in addition, there is a PFCT construction. In Georgian, however, there is an almost perfect correlation between the choice of prefixed vs. non-prefixed forms and the choice of the other tense-aspect categories, in such a way that the Present (Indicative and Subjunctive) and Imperfect forms are practically always without prefixes and the Aorist, Future, Conditional, Perfect and Pluperfect are normally prefixed. In the questionnaire, there are only two or three counterexamples to these correlations, all with the Future. It thus seems that what has happened here is a wide-reaching integration of the prefixing system with the 'ordinary' TMA-system.

Although the possible historical connections are rather unclear, Georgian, Hungarian and the Slavic languages form a contiguous geographical area. If we want to make judgements about the universality of the 'Slavic' system of aspects, it would be nice to have information from other areas about similar systems. In our questionnaire material, no such examples seem to exist: however, there is in the literature one notable example outside the area talked about

above, viz. Margi, which is a Chad language described in Hoffmann 1963. Since this is not a widely known fact, it may be a good idea to give some details about the Margi TMA system here.

According to Hoffmann, there is in the Margi verb an opposition between a perfective and an imperfective aspect. This is linked up with a derivational process, by which suffixes are added to verb stems. These suffixes seem to have a function very similar to that of verb prefixes in Slavic and Georgian, as Hoffmann notes (116). The most important suffixes are: *-ani*, 'causative'; *-ari*, '(the action is done) only a bit'; *-bá*, 'out'; *-la*, 'down'; *-na*, 'away, off'; *-nya* '(to consume) until nothing is left'; *-nggəri*, 'on top'; *-ri* (varying interpretation), *-wá*, 'into'. For transitive verbs, 'it can be assumed that the simple verb stem is imperfective, while the derivative stems generally speaking are perfective. Intransitive verbs apparently can be used also perfectly in their simple stems, but sometimes derivatives are used, it seems, to stress the fact of completion.'

In addition to this system, there are also inflectional and periphrastic TMA categories. There is a category called 'Progressive', marked by a particle *švər*, which is used for actions in progress but also for habitual action in the past – something which is not very common for PROGRESSIVE and which makes it look a bit more like a general Imperfective. In non-progressive forms, there is a distinction between Present and Past (although it is unclear how well this corresponds to the usual present-past distinction). In addition, there is a Narrative (marked by a particle *gà*) and an Aorist, which seems to be an alternative to the non-progressive Past for a few verbs of motion. Regrettably, Hoffmann is not always explicit concerning the relation between the derivational aspectual system and the other categories, except for the following statement concerning the past tense:

While the past tense of a perfective verb stem means 'somebody actually did (and finished) the action', the past tense of an imperfective verb only means 'somebody started to do the action (but did not finish it)' or 'somebody tried to do the action (but did not succeed)'. Thus the past tense of an imperfective verb resembles the Latin *imperfectum de conatu* fairly closely ... Also in another case the difference appears clearly. When a second verb is joined to a previous predicate with *gà*, the latter must be translated by 'in order to (do that action)', where the verb is imperfective, but by 'and (did that action)', where the verb is perfective... (115)

It appears from this quotation that in spite of the formal similarities between the Slavic-Georgian and the Margi systems, there are quite significant differences in the use of the categories. An Imperfective

Past in e.g. Russian is primarily used for on-going and habitual actions in the past: here, Margi apparently uses the Progressive. (Hoffmann does not indicate if this form can be formed from both aspects or not; the examples seem to be derived from imperfective verbs, however.) As for the uses mentioned by Hoffmann for the Past of Imperfective verbs in Margi, only the 'conative' is normally found with Imperfective Pasts in other languages, and, in fact, it is not quite clear if the Margi cases are really parallel to e.g. the Latin ones. One possible interpretation of the facts is that what Hoffmann calls Past is rather an example of PERFECTIVE, which applied to an imperfective verb, receives an inceptive or a conative interpretation. To sum up, the aspect systems that employ counterparts to Slavic perfectivizing prefixes seem to exhibit significant variations both in their semantics and in the interaction of the perfectivizing process with other parts of the TMA system. One possibility is that there is a general diachronic process by which morphemes with a semantics like the Slavic verbal prefixes are 'grammaticalized' as markers of 'perfectivity', but that these tend to be gradually integrated into the TMA system so as to finally lose their autonomous role, as seems to have almost happened in Georgian. In the East and West Slavic languages this tendency may have been counteracted by the development of imperfectivizing processes, such as the 'Iteratives', and by the fact that the old Aorist:non-Aorist opposition had disappeared. As to the identity or non-identity of the Slavic Perfective:Imperfective opposition with the cross-linguistic PFV:IPFV, it might be suggested that the question is wrongly put, since the categories in fact belong to different levels in the grammar: the Slavic-type aspect categories are no proper 'inflectional categories' but rather 'derivational' or maybe even 'grammaticalized lexical categories'. The semantic differences that we have seen might then reflect the fact that the Slavic categories grammaticalize 'perfectivity:imperfectivity' on the lexical or derivational level rather than in inflectional morphology. It is interesting to note that Finnish and Estonian, the languages in the sample where perfectivity is involved in quite another part of the grammar – (case marking), in several respects behave like the Slavic categories, as can be seen from table 3.2.

Hopper & Thompson 1980 have argued that perfectivity is one of the factors in what they call 'transitivity', a cluster concept which, in addition to transitivity in the usual sense, also involves a number of other things such as referential status, dynamicity, foregrounding etc. (I discuss the application of this idea to Finnish in Dahl (forthcoming).) I suspect that it will turn out that the involvement of perfectivity in the transitivity system of a language will be much clearer in systems of the Slavic or Finno-Ugric type than elsewhere: that is, the more inflectional PFV:IPFV categories will have less to do with transitivity.

The progressive (PROG)

The category PROG is found in at least 28 languages in the material (see table 3.7), and in one or other variety it is or may be represented in several others. The percentage in the small sample is slightly lower: there, unquestionable PROGs are found in five languages (27 per cent). The category seems to be more frequent among the Indo-European languages, which at least partly explains this discrepancy. The average frequency in the questionnaire of the clear PROGs is 24. The total frequency of those categories is 653, which makes it a rather well represented category.

Table 3.7 PROG categories

Language	Code	Description	Frequency	C	HR
PROG					
Chinese (Man.)	Z=	Particle <i>zài</i>	17	.71	1.0
Indonesian	SE	<i>sedang</i> + V	17	.66	.88
Sundanese	N=	<i>nuju</i> + V	17	.66	.88
Işekiri	WiG	<i>wínrǎǎbá</i> + V	16	.65	.88
Maltese	Q	<i>qiegħed</i> + i	19	.64	.84
Persian	D	Aux. <i>dastan</i> + V	26	.64	.85
Swedish	X	<i>hålla på att</i> + Inf.	16	.62	.81
English	BG	<i>be</i> + Gerund	39	.62	.82
Tamil	Z	PrP + Aux. <i>irykka</i>	29	.61	.79
Akan	pr	Progressive	23	.61	.78
Limouzi	T	<i>être en train de</i> + Inf.	22	.60	.77
Javanese	L	Aux. <i>lagi</i> + V	22	.60	.77
Spanish	P	Aux. <i>estar</i> + PrP	37	.60	.78
French	T	<i>être en train de</i> + Inf.	12	.59	.75
German	G	Particle <i>gerade</i>	7	.58	.71
Thai	K=	Aux. <i>kamlan</i> + V	18	.57	.72
Bengali	T	PrP+(contracted) copula	38	.57	.74
Portuguese	E	<i>estar</i> + <i>a</i> + Inf.	24	.56	.71
Maori	a	<i>e</i> + V + <i>ana</i>	35	.56	.71
Italian	S	Aux. <i>stare</i> + gerundio	16	.55	.69
Bandjalang	a	Cont./Antipassive	50	.54	.72
Yoruba	pr	<i>n</i> + V	21	.50	.62
Thai	Y	V + Aux. <i>yuú</i>	20	.49	.60
Fitzr. Cr. Kriol	d	Suffix <i>-bat</i>	41	.49	.63
Hindi/Urdu	R	Inf. + PtP of <i>rahna</i> + Cop.	31	.49	.61
Maori	c	<i>te</i> + V	19	.48	.58
Limouzi	Ea	<i>être a</i> + Inf.	9	.47	.56
Sotho	Nt	Aux. <i>ntse</i> + dep. form	12	.43	.50

Table 3.7 continued

<PROG					
Catalan	P	Aux. <i>estar</i> + PrP	11	.32	.36
Punjabi	R	V + PtP of <i>rahna</i> + <i>hona</i>	21	.31	.38
Guarani	e	Emphatic	10	.06	.10
Finnish	OX	<i>on</i> + 3rd Inf. in inessive	7	.03	.00

There is quite a strong tendency for PROG to be marked periphrastically (about 85 per cent of the cases), most often by auxiliary constructions.

Table 3.8 Distribution of PROG (n>5)

Rank no.	No. of categories	Examples
1	26	831
2	24	51
3	23	61
4	22	91 101 111
7	21	71 121 1551
10	20	81 241 861 1562
14	17	1402
15	16	161 841 1951
18	15	171
19	14	852 862
21	13	281
22	11	1611 1811
24	10	1481 1661 1711 1761
28	9	1861
29	7	851 961
31	6	1961
32	5	131 141 282 981

Semantics of PROG Table 3.8 shows the distribution of PROG in the questionnaire. The prototypical uses of PROG are listed in table 3.9. These all involve what could be labelled an 'on-going activity'. 'To go on' is basically a relation between a dynamic situation and a point in time. Accordingly, PROG is normally not used for stative constructions. The label 'durative' for PROG, which is quite commonly found in the literature, is misleading in that it gives the impression that PROG is used in contexts where the duration of a process is stressed. As we just noted, PROG naturally occurs with punctual temporal reference as in (3.15). In a context like (3.16), it is avoided.

(3.15) At twelve o'clock sharp, John was still writing the letter

(3.16) ?John was singing/sang for ten minutes

Table 3.9 Prototypical occurrences of PROG

Sent. 83 Verb 1 No. of categories: 29

Father to child:

(Please do not disturb me), I WRITE a letter

Sent. 6 Verb 1 No. of categories: 27

Q: What your brother DO right now? (=What activity is he engaged in?) A by someone who can see him:

He WRITE a letter

He WRITE a letter

Sent. 5 Verb 1 No. of categories: 27

Q: What your brother DO right now? (=What activity is he engaged in?) A by someone who can see him:

He WRITE letters

He WRITE letters

Sent. 11 Verb 1 No. of categories: 27

A: I talked to my brother on the phone yesterday. B: What he DO?

(=What activity was he engaged in?)

He WRITE letters

Sent. 9 Verb 1 No. of categories: 26

A: I went to see my brother yesterday. B: What he DO? (=What activity was he engaged in?)

He WRITE letters

Sent. 12 Verb 1 No. of categories: 26

A: I talked to my brother on the phone yesterday. B: What he DO?

(=What activity was he engaged in?)

He WRITE a letter

Sent. 10 Verb 1 No. of categories: 26

A: I went to see my brother yesterday. B: What he DO? (=What activity was he engaged in?)

He WRITE a letter

In languages with a perfective-imperfective distinction, the prototypical PROG contexts would be imperfective. In fact, in some treatments, PROG and imperfective aspect are partially or wholly identified: thus, Friedrich 1974 lumps them together under the heading 'durative', and also Comrie 1976 speaks of progressivity as a special case of imperfectivity. It may therefore be worth emphasizing that there are clear features that distinguish PROG from imperfective aspect, viz:

- (i) In contradistinction to PFV:IPFV, which is strongly correlated with the distinction between past and non-past time reference, PROG is usually independent or almost independent of time

reference – in other words, it is used both of the present, the past and the future (although less frequently of the latter, cf. the lower rating of (Q.16)). If there are restrictions, it is rather the present than the past that is favoured with progressives (see discussion below).

- (ii) PROG is quite infrequently extended to habitual meaning – there are only a handful of PROG markings in e.g. (Q.18–21).
 (iii) As noted above, PROG is normally used only of dynamic – that is, non-stative – situations. A typical example would be (Q.118), which contains the verb *know* – not a single PROG marking is found in the material for this sentence.

We note also that PROG and PFV:IPFV are very different in the ways in which they are marked – PROG, as we noted above, is rather consistently marked periphrastically, whereas – as discussed above – the marking of PFV:IPFV is usually much less straightforward.

It should be pointed out that the cases which distinguish IPFV and PROG are relatively restricted in frequency, and that it is to be expected that diachronically, a category can shift from IPFV to PROG or vice versa. In fact, the latter development is attested in Hindi/Urdu and Punjabi, where an older PROG construction (Present Participle + Copula) has been extended to a general IPFV – basically by conquering the territory of the old Simple Present, and has then lost its focal uses to a new PROG construction (Present Participle + Past Participle of the verb *rahna* + Copula).

When a PROG is opposed to a 'Simple Present' there is also the possibility of a shift to a situation where PROG takes on the role of an IPFV and the earlier 'Simple Present' develops into a category with primarily habitual use. This is apparently taking place in Turkish and Azerbaijani.

Extended uses of PROG

The English Progressive has a number of secondary uses, such as what Comrie 1976 calls the 'contingent state' use, that is, to refer to a temporary habit or state, as in (3.17):

(3.17) I'm living in London (i.e. for the time being)

(Q.72) was included in the questionnaire in order to elicit parallels in other languages:

(Q.72)

Context: This week I have to go to work early

Sentence: I RISE at six in the morning (alt: at dawn)

The result was negative: not a single instance of PROG was recorded outside of English. However, many informants clearly did not understand the sentence in the intended way, and the data are not

too reliable. Still, the result is consistent with Comrie's remark (1976, 32) that the English Progressive 'has, in comparison with progressive forms in many other languages, an unusually wide range' – something which is corroborated by our statistics: the English Progressive ranks as third in frequency among the PROG categories in the material. We note here some of the cases of untypical uses of PROG in English: with verbs of saying (Q.109–11), 'reportive present' (Q.196–7) (somewhat unexpectedly perhaps), and perhaps most strikingly (Q.59), with the characteristic use of the Perfect Progressive. (For a detailed treatment of the secondary uses of the English Progressive, see Ljung 1980.)

Problematic cases

We said above that PROG hardly ever extends to habitual meaning. There is at least one category where progressive and habitual uses are combined and where the analysis is problematic, viz. the Fitzroy Crossing *-bat* form: it has a fairly high correlation with PROG, but deviates from the typical PROG in several respects and may be better analysed as IPFV for non-statives. The Quechua 'Progressive' combines progressive and stative uses.

Time-restricted progressives

There are some exceptions to the generalization that progressives can be used equally well with present and with past time reference. Progressive constructions restricted to present time reference are found in five languages, listed in table 3.10. There is only one example of a PROG-PAST among the major categories, viz. the Wolof *dafa don* construction. Although the material is quite limited, the tendency thus seems to be that if there is a restriction on temporal reference in a progressive, it is to the present. It is an open question whether PRES-PROG should be treated as a variety of PROG or a category in its own right. We may note, however, that the predominant trend

Table 3.10 PRES-PROG categories

Language	Code	Description	Frequency
Hawaiian	pr	<i>ke + V + nei</i>	25
Karaboro	C	<i>ma + V</i>	10
Kikuyu	ra=	Prefix <i>ra-</i> : suff 0	43
Oromo (Galla)	aG	Imperfetto + aux. <i>gira</i>	11
Tigrinya	pr	Progressive	7

for restrictions on time reference to favour the present marks off PROG both from PFV and from the habitual categories dealt with in chapter 4.

The CONCLUSIVE (CONCL)

The assumption that there is a cross-linguistic category which can be labelled CONCLUSIVE (CONCL) is mainly based on some striking similarities between two language-specific categories, viz. the Japanese *-te shimau* construction and the Tamil construction with the auxiliary *vida*. There are eight examples in the questionnaire where both these constructions are used, viz. (Q.53–4), (Q.89), (Q.162) and the examples in the other narrative texts that correspond to (Q.162). Both constructions also seem to have the same basic meaning: 'to finish doing something', although the Japanese construction is said to have originally meant 'to close' (Martin 1975, 533). Martin also says that *-te shimau* can mean 'to do something completely'. Clearly, however, the present examples (as well as some of Martin's, some of which are quite similar to ours) involve an extension of this basic meaning: it is remarkable that many of them concern punctual actions ('step on', 'leave', and at least for Japanese 'die', e.g. (Q.165)), to which phasal operators such as 'finish' do not normally apply. Martin uses the label 'strong past' for some of his examples: perhaps 'strong perfective' might be a better term.

In (Q.162) and the parallel sentences in the other texts, one factor is the suddenness of the action, something which has clearly also influenced the choice of TMA category in other languages – in Sotho and Zulu, for instance, a so-called 'now-tense' may be used here (see p. 176). The suffix *-riataar-* in Greenlandic Eskimo is used here and elsewhere to signal a sudden event: due to its overlap with the CONCL categories, we have labeled it CONCL?, although this is more in order not to forget that there is a connection than as a well-founded hypothesis. In any case, constructions of the kind we have discussed in this section certainly deserve further study.

Habituals and generics

In this section, we shall treat a number of closely related categories, which we shall label as HAB, HABG and HABPAST. They have in common that they express actions that take place habitually or repeatedly. In addition, they may have a number of secondary uses.

All these categories have a rather low frequency in the questionnaire, and to get a fuller account of habitual categories we have in the tables included some examples within parentheses that do not qualify

as 'major TMA categories' according to the frequency criterion. The tables that give the distribution of habitual categories also have been made up with a more generous definition of which categories should be included.

The HABITUAL (HAB)

Table 3.11 HAB categories

Language	Code	Description	Frequency	C	HR
HAB					
Guarani	A	Suffix <i>-akostumbrakuri</i>	4	.71	1.00
German	P	<i>pflügen</i> + Inf.	4	.71	1.00
Swedish	y	<i>bruka</i> + Inf.	5	.71	1.00
Georgian	X	Particle <i>xolme</i>	12	.67	.92
Kammu	K	<i>ku</i> + V	9	.61	.78
Czech	it	Iterative verb	9	.55	.67
Akan	T=	<i>-taa</i> + V	8	.52	.63
Wolof	FD21	<i>dan</i> + <i>de</i> + V	7	.49	.59
<HAB					
Yoruba	Mpr	<i>ma</i> + <i>n</i> + V	6	.30	.33

Table 3.12 Distribution of HAB ($n > 1$)

Rank no.	No. of categories	Examples
1	10	181 191
3	8	201 211
5	4	311 711 1931
8	3	251 401 1911
11	2	1022 1192 1921

This category meets the requirement for being a 'major TMA category' in four languages, which are listed in table 3.11. In addition, there are several languages in which (optional) categories are attested whose distribution agrees with the HAB 'ranking list' but which have too low a frequency to qualify as 'major TMA categories'. It is fairly plausible that similar categories can be found in other languages although they have not found their way into the questionnaires. There are too few instances of HAB to make clear statements about marking types: if we count all the categories in table 3.11, the majority are periphrastic, however.

It may be easiest to characterize HAB by what it is not. A term commonly applied to HAB categories is 'iterative'; this suggests that 'iteration' is a relevant criterion for their use. If we take 'iteration' to mean that something happens more than once, the suggestion is clearly false: the difference between 'once' and 'twice' or even 'seven times' is almost totally irrelevant to HAB, as is witnessed by the fact that it is hardly used at all in sentences (Q.92-7). Neither is HAB used in generic sentences, like (Q.73) – this is what distinguishes it from the category we call HABG. It appears that the cases where HAB is typically used are those in which the adverb *usually* is possible in English (see tables 3.12-13). These involve quantification over a set of occasions which is given explicitly or by context. For instance, in (Q.20), the set of occasions consists of the time intervals immediately following upon the breakfasts the speaker's brother had last summer. The use of HAB indicates that what is expressed in the sentence took place in the majority of those occasions. Such sentences differ from generic ones by their lack of lawlikeness. Notice that those sentences which concern personal habits but are more like generics (Q.25, 71) have significantly less frequent use of HAB.

Table 3.13 Prototypical occurrences of HAB

Sent. 19 Verb 1 No. of categories: 10

Q: What your brother usually DO after breakfast? A:
He WRITE a letter

Sent. 18 Verb 1 No. of categories: 10

Q: What your brother usually DO after breakfast? A:
He WRITE letters

Sent. 21 Verb 1 No. of categories: 8

Q: What your brother usually DO after breakfast last summer? A:
He WRITE a letter

Sent. 20 Verb 1 No. of categories: 8

Q: What your brother usually DO after breakfast last summer? A:
He WRITE letters

Sent. 31 Verb 1 No. of categories: 4

Of a visible lake, what the water is usually like
It BE COLD

Sent. 193 Verb 1 No. of categories: 4

I'll tell you what happened to me sometimes in the forest
I SEE a snake

Table 3.13 continued

Sent. 71 Verb 1 No. of categories: 4
Talking about the speaker's habits: I like to be up early.
I RISE at six in the morning (alt: at dawn)

Sent. 40 Verb 1 No. of categories: 3
Q: Do you know my brother?
(Yes,) I MEET him often (up to now)

The HABITUAL-GENERIC (HABG)

Table 3.14 HABG categories

Language	Code	Description	Frequency	C	HR
HABG					
Isekiri	K	ka + V	22	.65	.86
Wolof	FD	Aux. <i>dafa</i> + <i>de</i> + V	18	.61	.78
Greenl. Eskimo	t=	Suffix <i>-tar-</i>	28	.60	.79
Maori	h	V + <i>ai</i>	8	.60	.75
Number of categories: 4. Total frequency: 76.					
<HABG					
Cebuano	u=v	<i>mag</i> + R	20	.45	.00
Seneca	I	Iterative aspect	18	.37	.44

Table 3.15 Distribution of HABG (n>2)

Rank no.	No. of categories	Examples
1	6	181 191 731
4	5	201 211 711 741 1911
9	4	251 401 441 752 1022 1931

Clear examples of HABG are found in four languages in the total sample (see table 3.14). The average frequency in the questionnaire is 25. HABG seems to be obligatory except in Maori (dialect differences?) and possibly Wolof. Among the HABG categories, there are examples of most marking types: the total number is too small for it to be possible to discern any clear tendencies.

Table 3.16 Prototypical occurrences of HABG

Sent. 19 Verb 1 No. of categories: 6
Q: What your brother usually DO after breakfast? A:
He WRITE a letter

Sent. 18 Verb 1 No. of categories: 6
Q: What your brother usually DO after breakfast? A:
He WRITE letters

Sent. 73 Verb 1 No. of categories: 6
Q: What kind of sound do cats make?
They MEOW

Sent. 71 Verb 1 No. of categories: 5
Talking about the speaker's habits: I like to be up early.
I RISE at six in the morning (alt: at dawn)

Sent. 191 Verb 1 No. of categories: 5
I'll tell you what happens to me sometimes when I am walking in the forest
I SEE a snake

Sent. 74 Verb 1 No. of categories: 5
Q: What do your cats do when they are hungry?
They MEOW

Sent. 20 Verb 1 No. of categories: 5
Q: What your brother usually DO after breakfast last summer? A:
He WRITE letters

Sent. 21 Verb 1 No. of categories: 5
Q: What your brother usually DO after breakfast last summer? A:
He WRITE a letter

Somewhat deviant examples (labelled <HABG) are found in Cebuano and Seneca (see further discussion below).

HABG differs from HAB primarily by being used also in generic sentences (see tables 3.15–16). It is possible that it should be considered a variety of HAB and that the generic cases are a secondary use. From our material it seems, though, that both the 'habitual' and the 'generic' uses are equally well established.

The semantics of generic sentences has been discussed in a large number of works in recent years (see e.g. Dahl 1975, Carlson 1977, 1979, 1982). The characteristic property of those sentences is their 'nomic' or 'lawlike' character – they describe the typical or characteristic properties of a species, a kind, or an individual. It seems to be rare for a language to have an overt and unequivocal TMA marking of

a sentence as being generic. The most frequent case is for generic sentences to be expressed with the most unmarked TMA category, as in English, where the Simple Present is used. Indeed, 'generic' sentences seem to be the cases that are most often unmarked for TMA. It seems again, however, that these unmarked categories almost always have other uses.

'Generic' sentences are often claimed to be timeless. However, as I pointed out in Dahl 1975, a law or lawlike statement may be restricted in time, and this time may well be in the past or in the future. The exact status of sentences like (3.18) has been disputed; we did not include any such example in the questionnaire, since it was hard to find one which would not cause translation problems. It is to be expected that those languages which mark past time reference in habitual sentences will do so also in generic sentences.

(3.18) Dinosaurs ate kelp

The <HABG categories in Seneca and Cebuano behave peculiarly in several respects – they are not used in (Q.20–1), but as a compensation they occur in (Q.93–6) (the cases discussed above of true 'iteration'). (Also the Greenlandic Eskimo HABG category is used in the latter.)

Marking of time reference in HAB and HABG

Of the languages with HAB, at least the following mark past habituais with a combination of HAB and a standard past time marking: Akan, Czech, Guarani, German, Georgian, Swedish, Hungarian.

Combinations of HAB and FUT are more difficult to find in our material, which may be due to lack of suitable examples.

A combination of HABG and a past tense is found in Wolof; a combination of HAB and FUT in Isekiri.

The HABITUAL-PAST (HABPAST)

HABPAST is used as a label for categories which are mainly used for habitual sentences with past time reference and are not analysable as consisting of HAB or HABG combined with a regular past tense. Counting also categories with a frequency below 6 in the questionnaire, HABPAST appears in 6–7 languages in the total sample and in 4 in the small sample (22 per cent) – see table 3.16. There is some concentration in certain genetic groups. In view of the low text frequency of this category, the generalizations below should be taken with some caution.

Table 3.17 HABPAST categories

Language	Code	Description	Frequency	C	HR
HABPAST					
Bandjalang	h	Habitual Past	4	.71	1.00
English	U	<i>used to</i> + V	5	.62	.80
Seneca	R	Repetitive particle	13	.60	.77
Alawa	g	Suffix <i>-gay</i>	8	.60	.75
Oneida	21	Serial Past	4	.44	.50
HABPASTc					
Azerbaijani	32P	32 + P	5	.62	.80
Bengali	25	Past Frequentative	13	.33	.38

Table 3.18 Distribution of HABPAST (n>2)

Rank no.	No. of categories	Examples
1	6	201 211 1022
4	4	261
5	3	441

Table 3.19 Prototypical occurrences of HABPAST

Sent. 102 Verb 2 No. of categories: 6
The boy used to receive a sum of money now and then
When the boy GET the money, he BUY a present for the girl

Sent. 20 Verb 1 No. of categories: 6
Q: What your brother usually DO after breakfast last summer? A:
He WRITE letters

Sent. 21 Verb 1 No. of categories: 6
Q: What your brother usually DO after breakfast last summer? A:
He WRITE a letter

The basic semantics of HABPAST appears to be describable as a combination of HAB and past time reference. For its distribution and prototypical cases, see tables 3.18–19. It is not quite clear if

HABPAST can be used in 'past time generics': at least the English example (3.19) sounds rather strange:

(3.19) Dinosaurs used to eat kelp

There may be a connection between HABPAST and remote past time reference: such a connection is argued for by Kučera 1981 for the Czech HAB ('iterative') in the past tense.

At least two languages, Bengali and Azerbaijani, exhibit a secondary use of HABPAST that should be noted, viz. in counterfactual sentences (e.g. Q.106).

There seems to be a predominance for morphological marking of HABPAST – exceptions are English and Seneca.

Categories combining habitual/generic uses with others Since most languages do not have separate habitual or generic categories, it follows that categories that combine habitual and/or generic uses with others are fairly common. In most cases, as noted elsewhere, this tends to be either the most unmarked category ('Simple Present' in languages with a past–non-past distinction) or some imperfective category. Leaving these cases alone, there are a small number of 'marked' categories that combine the domain of use of HAB with other uses, commonly connected with future time reference (see chapter 4 for discussion).

4

Tense categories

Under the heading 'Tense categories' we treat those categories in the semantics of which time reference seems to be the primary dimension, and also some others which are closely related to the former and cannot be naturally subsumed elsewhere.

Future time reference

Before going into a discussion of the ways in which languages refer to the future, let us note that the future differs epistemologically – and maybe also ontologically – from the present and the past, as Aristotle noted. We cannot perceive the future directly or 'remember' it – at least, this is what conventional wisdom tells us. We shall not try to answer the metaphysical question whether the future is determined or not, or the corresponding truth-theoretical question, whether propositions about the future have determinate truth-values or not. It is clear, however, that the special status of the future will have linguistic consequences, in particular as regards the semantics of TMA systems.

Normally, when we talk about the future, we are either talking about someone's plans, intentions or obligations, or we are making a prediction or extrapolation from the present state of the world. As a direct consequence, a sentence which refers to the future will almost always differ also modally from a sentence with non-future time reference. This is the reason why the distinction between tense and mood becomes blurred when it comes to the future. This has been pointed out again and again in the literature, but I hope to be able to give the problem a slightly new twist below in discussing the semantics of FUT, the universal category I assume underlies most of the forms called 'future tenses'.

FUT

FUT appears in around 50 of the languages in the sample (see table 4.1). This means that the number of languages that do not have this

Table 4.1 FUT categories

Language	Code	Description	Frequency	C	HR
FUT					
Georgian	30	Future	41	.64	.85
Quechua	3=	Future (suffix <i>-ta</i>)	12	.64	.83
Sotho	Tl	Aux. <i>ila</i>	17	.63	.82
Bugis Makassar	E	<i>ero</i> + V	11	.63	.82
Afrikaans	S1	<i>sal</i> + V	23	.63	.83
Bengali	30	Future	29	.63	.83
Bandjalang	3	Future	34	.62	.82
Alawa	30	Future	34	.62	.82
Italian	30	Future	27	.62	.81
Latin	31	Futurum	21	.62	.81
Hindi/Urdu	3=	Future	30	.61	.80
Portuguese	I	<i>ir</i> + <i>a</i> + Inf. of V	9	.61	.78
Seneca	Pf	Punctual future	32	.60	.78
Karaboro	G	<i>ga</i> + V	26	.60	.77
Guarani	30	Future	12	.59	.75
Hungarian	F	<i>fog-</i> + SM + Inf.	12	.59	.75
Hebrew	30	Future	30	.59	.77
Wolof	Y=	<i>dina</i> + V	16	.59	.75
English	W	<i>will</i> + Inf.	20	.59	.75
Bulgarian	30	Future	20	.59	.75
Kammu	C	<i>cee</i> + V	24	.59	.75
Greek (Modern)	T	Particle <i>tha</i>	23	.58	.74
Romanian	O	Aux. <i>o</i> + Subj.	14	.57	.71
Oneida	3	Future	25	.57	.72
Akan	3=	Future	16	.55	.69
Maltese	S	<i>ser</i> + Imperfective	6	.55	.67
Fitzr. Cr. Kriol	G	Aux. <i>garra</i> + V	12	.54	.67
German	W=	Aux. <i>werden</i> + Inf.	17	.53	.65
Limouzi	30	Future	17	.53	.65
Catalan	30	Future	17	.53	.65
Punjabi	30	Future	30	.53	.67
Thai	C=	Aux. <i>cà</i> + V	26	.52	.65
French	30	Futur Simple	16	.51	.63
Spanish	30	Future	10	.50	.60
Greenl. Eskimo	30	R + Suffix <i>saa</i>	15	.50	.60
Tigrinya	30	Future	12	.49	.58
Swedish	S	Aux. <i>skola</i> + Inf.	12	.49	.58
Romanian	V	Aux. <i>voi</i> + Inf.	14	.48	.57
Kurdish	30	Future	27	.48	.59
Javanese	A	<i>arep</i> + V	9	.47	.56
Persian	X	Aux. <i>xastan</i> + Inf.	16	.47	.56
Işekiri	W	<i>waá</i> + V	36	.46	.58
Turkish	30	Future	13	.46	.54
Beja	Y	Imperative + <i>-y-dy</i>	6	.44	.50
Yoruba	Y	<i>yio</i> + V	8	.44	.50

Table 4.1 continued

Azerbaijani	31	Categorical Future	12	.43	.50
Arabic (ModSt)	3	Future	13	.40	.46
Indonesian	A	<i>akan</i> + V	13	.40	.46
<FUT					
Portuguese	3=	Future	8	.33	.38
Yoruba	M=	<i>ma</i> + V	6	.30	.33
Chinese (Man.)	H	Particle <i>hui</i>	11	.23	.27
Oneida	c	Continuative	9	.19	.22
Spanish	F	Aux. <i>ir</i> + <i>a</i> + Inf.	5	.18	.20
FUTi					
Czech	B	<i>budu</i> + Inf.	9		
Russian	B	<i>budu</i> + Inf.	8		
Polish	3=	Aux. <i>być</i> + Past/Inf.	13		

category is fairly small. It therefore makes more sense to discuss them as a group than the languages where FUT is manifested – see the section called ‘Languages without FUT’ below. The average frequency of FUT in the questionnaire is 16.7 (8 per cent).

Among the FUT categories in our material, 27 – more than 50 per cent – are marked morphologically. This puts FUT among the three categories that are most often marked morphologically.

The morphemes that mark FUT are often used in combination with others to form complex TMA categories, such as conditionals. This gives rise to problems of the kind discussed on p. 67. Since conditionals do not in general meet the conditions for being regarded as ‘major TMA categories’, we have not treated them separately here. The statistics concerning the use of FUT therefore also includes FUT markers when used in complex categories, which may sometimes have distorted the figures somewhat.

The so-called Future tense in Tamil has not been labelled FUT, since its correlation values are too low. The reason for this is its extensive use in habitual contexts. For some other examples of categories which are used both with future time reference and in habitual contexts, see the section on PRED below.

Semantics of FUT It seems fairly clear from the data presented in tables 4.2-3 that the most typical uses of FUT involve actions that are planned by the agent of the sentence (e.g. Q.152:2, 103:2, 104:2). Cases of ‘pure prediction’ (e.g. Q.36) get much lower frequencies. This suggests that ‘intention’ is part of the prototype of FUT, something which seems to be in agreement with what we said above about the modal element of future tenses. It would be wrong, however, to suggest that ‘intention’ is the essential semantic feature of

Table 4.2 Distribution of FUT

Rank no.	No. of categories	Examples
1	42	1032 1522
3	38	271
4	37	1042
5	36	151
6	35	221 361
8	34	231
9	33	171
10	32	161
11	31	1452
12	28	792
13	27	811 1463
15	23	1072 1142
17	21	802 1252
19	19	782
20	18	2051
22	17	772
23	15	1521
24	14	721 1052
26	13	1081
27	12	752 1071
29	11	1031 1451
31	10	771 1062
33	9	781 801 1041
36	8	241
37	7	751 791 1242
40	6	1462
41	5	1162 1313 1461
44	4	1152 1312

FUT. Notice, on one hand, that 'intention' more often than not is no necessary condition on the use of FUT: in the majority of cases, FUT can also be used for cases where the intentional element is lacking. If we have a sentence expressing intention but with no element of prediction, i.e. where the speaker takes no stand on whether the action will take place or not, it does not generally seem possible to use FUT. There is no such sentence in the questionnaire, but at least for a language like English, a sentence such as *He will go* cannot be used if the speaker doesn't believe that the going will actually take place. Cf. *He intends to go*, which has no such implication.

We thus see that 'future time reference' is a more constant element of FUT – relatively speaking! – than the modal features of this category, which in most cases may or may not be present. 'Future time reference' could therefore be regarded as a dominant feature of FUT

Table 4.3 Prototypical occurrences of FUT

Sent. 152 Verb 2 No. of categories: 42 Said by a young man When I GROW old, I BUY a big house
Sent. 103 Verb 2 No. of categories: 42 The boy is expecting a sum of money When the boy GET the money, he BUY a present for the girl
Sent. 27 Verb 1 No. of categories: 38 A: My brother has got a new job at an office. He will start tomorrow. B: What kind of work he DO there? He WRITE letters
Sent. 104 Verb 2 No. of categories: 37 The boy thinks that he will perhaps get a sum of money If the boy GET the money, he BUY a present for the girl
Sent. 15 Verb 1 No. of categories: 36 Q: What your brother DO if you don't go to see him today, do you think? A: He WRITE a letter (to me)
Sent. 36 Verb 1 No. of categories: 35 (It's no use trying to swim when we get to the lake tomorrow.) The water BE COLD (then)
Sent. 22 Verb 1 No. of categories: 35 Q: What are you planning to do right now? A: I WRITE letters
Sent. 23 Verb 1 No. of categories: 34 Q: What are you planning to do right now? A: I WRITE a letter

in the sense discussed in chapter 1. The traditional view of the Future as a tense can thus be defended.

Some FUT categories cannot normally be used for 'pure prediction'. The Swedish *skall* construction is an example in point. Because of this, most modern grammars of Swedish do not call *skall* a future auxiliary. In spite of this, it has quite a high correlation with the hypothetical universal category – $C = .58$,¹ a rather startling fact, in particular in comparison to the alleged 'pure future' auxiliary *kommer att*, which has the C-value .38.

Another interesting observation about *skall* is that the constraint against using *skall* for 'pure predictions' holds only for the Present tense of the auxiliary. The Past tense, *skulle*, can in fact be used for

'predictions in the past'. To illustrate this, consider a sentence such as (4.1):

(4.1) Det kommer/blir/*skall bli varmt i eftermiddag 'It'll be hot this afternoon'

If we embed (4.1) as the complement of a verb of saying in the Past, we get:

(4.2) Pelle sa, att det *kom att bli/*blir/skulle bli varmt på eftermiddagen 'Pelle said that it would be hot in the afternoon'

In (4.1), we use either the *kommer att* construction or the Present. In (4.2), neither of these is possible. This fact seems rather arbitrary, but can be accounted for in the prototype theory if we assume that 'intention' is a secondary feature of FUT which is relevant for *skall* in default cases but is neutralized in contexts where the Past of *skall* is called for.

Strictly speaking, 'intention' is too narrow a criterion to cover even the non-Past of *skall*, since it can be used also for e.g. obligations and in various other modal senses. Also, interestingly, the addition of an adverb such as *nog* seems to make it much easier to use *skall* without the implication of someone's intention:

(4.3) Det här skall nog gå bra 'This will probably go well'

The prototypical uses of FUT that we have cited have clear future time reference. It is well known that many categories called 'Future' can also be used for 'predictions' or 'inferences' about non-future states of affairs, as in the English (4.4):

(4.4) That will be the postman

It is reasonable, however, to regard this as a secondary use of FUT categories, which preserves one element of the prototypical uses, viz. that of prediction (in the wide sense of that word) although leaving out the others. A consequence of this is that it is not possible to attribute any absolute status to the dominance of 'future time reference' in the semantics of FUT.

Summing up, we see that the semantics of FUT can be best described in terms of a prototype involving at least the three features 'intention', 'prediction', and 'future time reference'.

Languages without FUT

The following languages lack a category which can be identified as FUT or <FUT in our material: Tunisian Arabic, Estonian, Finnish, Japanese, Kikuyu, Sundanese, Maori. (Only one of these is in the small sample!) However, in all of these except Estonian and Finnish, there are categories which are labelled as Future tenses in traditional

descriptions or seem to be used primarily with future time reference, although their frequency is too low in the questionnaire to allow certain identification – that is, they do not occur more than five times, which is less than a third of the average frequency of FUT.

In view of the rarity of FUT-less languages, it may be of some interest to look at the ways in which future time reference is expressed in one such language, viz. Estonian. In that language, most of the sentences where FUT is normally used, even the most prototypical ones, employ the Present tense:

(Q.152:ES) Kui saan vanaks, ostan suure maja 'When I grow old, I'll buy a big house'

(Q.104:ES) Kui poiss raha kätte saab, ostab ta tüdrukule kingituse 'When the boy gets the money, he'll buy a present for the girl'

Notice, however, that the use of the verb *saan*, 'becomes', rather than *olen* in (Q.152) is in itself a signal of future time reference. Furthermore, since object marking partly depends on aspect in Estonian, present and future time reference is systematically distinguished in the transitive sentences with quantitatively delimited objects, such as 'the house' in (Q.152). In addition, Estonian has one extremely marginal construction, viz. the Present of *saa-*, 'become', followed by the so-called *-ma*-infinitive, that is used in predictive sentences. It seems to be restricted to impersonal copula constructions, however:

(4.5) Saab tore olema 'It will be fun'

Thus we see that even in Estonian, which otherwise appears to be one of the languages which cares least about the future, there are ways of signalling future time reference.

Obligatoriness of FUT

The general impression is that future time reference is less consistently marked than past time reference.² Such a statement is rather hard to make precise, though, since due to the epistemological differences between future and past time reference, it is not clear how one could find 'minimal pairs' to compare FUT and PAST. Some observations can be made, however. Among the majority group of languages that have categories that we have labelled as FUT, it is fairly common to find some variation between that category and others, in particular unmarked forms or present tenses. For instance, consider (Q.81):

(Q.81)

Context: Q: What HAPPEN if I eat this mushroom?

Sentence: You DIE

A non-future category may be used in (Q.81) in the following languages in the FUT group: Classical Arabic, Beja, Bugis Makassar, Catalan, German, Guarani, Italian, Maltese, Persian, Romanian, Spanish, Swedish, Tigrinya, Yoruba. In general, non-FUT forms seem to be possible above all with cases of 'pure prediction', but also in other situations, e.g. (Q.82), which is intended to be an example of what has been referred to as 'scheduling':

(Q.82)

Sentence: (According to the contract) we not WORK tomorrow

FUTi – FUT restricted to IPF

As has already been pointed out (chapter 3), in the East and West Slavic languages – Russian, Czech and Polish in our material – there is no morphological future tense except for the copula (*budet* in Russian). The future copula can be used to form a periphrastic future construction, but only of Imperfective verbs. FUTi is then the counterpart for the future of the much more frequently appearing PASTi.

Other categories which signal future time reference

Not all categories used for future time reference can be subsumed under FUT. It is fairly common that a language has more than one category in the future area. English has e.g. both *shall*, *will*, and *be going to*, as alternative future auxiliaries. Normally, such alternating constructions differ in more or less subtle ways in their semantics. It is reasonable to assume that one should be able to identify some cross-linguistically valid types here. Regrettably, the total number of examples in the questionnaire with future time reference is not large enough to make very reliable identifications of such types. The categories discussed below should therefore be regarded as somewhat tentative. They seem to have a rather peripheral status in the systems they belong to.

PRED PRED (acronym for 'predictive') is a category which is primarily used for future time reference which does not have an intentional element, that is, basically for making predictions. This is a rather rare category – depending on how we delimit it there are between three and five languages in the material that have it. The clearest examples are listed in table 4.4. We do not give any correlation coefficients, as the material is too small, but the sentences in the questionnaire where PRED is most often used are listed in tables 4.5-6. All PRED categories seem to be optional – there are always translation variants. They are all periphrastic.

In addition, habitual categories may be used secondarily in predictive contexts. The clear examples in our material are the two Turkic languages, where the so-called Aorist combines the uses of PRED and HAB (Azerbaijani) or HABG (Turkish).

Table 4.4 PRED categories

Language	Code	Description	Frequency
Azerbaijani	32=	Non-categorical Fut., Aorist	8
Japanese	D	<i>daroo</i> construction	13
Swedish	K	<i>kommer + att + Inf.</i>	9

Table 4.5 Distribution of PRED ($n > 1$)

Rank no.	No. of categories	Examples
1	4	811
2	3	161 171 361
5	2	151 752 1452

Table 4.6 Prototypical occurrences of PRED

Sent. 81 Verb 1 No. of categories: 4
Q: What HAPPEN if I eat this mushroom?
You DIE

Sent. 36 Verb 1 No. of categories: 3
(It's no use trying to swim when we get to the lake tomorrow.)
The water BE COLD (then)

Sent. 16 Verb 1 No. of categories: 3
Q: What your brother DO when we arrive, do you think? (What activity will he be engaged in?)
He WRITE letters

Sent. 17 Verb 1 No. of categories: 3
Q: What your brother DO when we arrive, do you think? (What activity will he be engaged in?)
He WRITE a letter

PROSP Comrie (1976, 64) calls constructions like the English *be going to* 'prospective'. A formally and semantically analogous construction to the English one is found in some Romance languages, e.g.

the French *aller* + infinitive, and in Afrikaans *gaan* + infinitive. In all those, verbs meaning 'to go' are employed. There are a few other categories in the material which have in common with the ones just mentioned that they are all used in (Q.22–3) but have low correlations to FUT:

(Q.22)

Context: Q: What are you planning to do right now? A:

Sentence: I WRITE letters

(Q.23)

Context: Q: What are you planning to do right now? A:

Sentence: I WRITE a letter

The only category whose frequency in the questionnaire is more than 5 is the Sundanese *bade* construction.

There is thus some evidence for postulating a cross-linguistic category PROSP. All the assumed examples of PROSP are periphrastic.

It should be noted that in some languages (e.g. Maori and Oromo), a PROG category may be used in (Q.22–3), as is also possible in similar cases in English.

Narrativity

Definitions of 'narrative discourse' and 'narrative context'

As in Dahl 1977 and other places, I define a **narrative discourse** as one where the speaker relates a series of real or fictive events in the order they are supposed to have taken place. As an example of a maximally short narrative discourse, Julius Caesar's famous statement (4.6) may be quoted:

(4.6) Veni, vidi, vici 'I came, I saw, I conquered'

In actual texts, such ideal or pure narrative discourses are of course relatively seldom found. Normally, the main story-line is continuously interrupted by various kinds of flashbacks and points of background information. This fact does not diminish the value of the concept of narrative discourse, though.

For the study of TMA systems, the most important concept here is that of a **narrative context**. We shall say that a sentence occurs in a narrative context if the temporal point of reference (in Reichenbach's sense) is determined by the point in time at which the last event related in the preceding context took place. Thus, the event referred to by *vidi* in (4.6) is understood to have taken place directly after that referred to by *veni*. Basically, this means that in a pure narrative

discourse, every sentence except the first one is in a narrative context. *Veni* in (4.6) is thus not in a narrative context – in spite of being in a narrative discourse. This makes sense, since the normal function of the first sentence of a narrative discourse is quite different from that of the others: it has to provide the temporal anchoring for the rest of the discourse by e.g. an explicit time adverbial (*one day last week, once upon a time* etc.).

TMA categories in narrative contexts

Narrative contexts behave in special ways with respect to TMA marking in many languages. A few languages have special narrative forms – these are discussed in the following section. More common than marking narrative contexts, however, is not marking them – quite a considerable number of languages use unmarked verb forms in narrative contexts. Even in languages that normally mark past time reference, e.g. English, it is often possible to use a so-called 'historical present' in narrative contexts. It would have been interesting to find out to what extent this is a universal phenomenon, but we decided it would have been too complicated to try and elicit historical presents from the informants. Some used it spontaneously, however, e.g. our Romanian informant (for texts B2, B4, B9).

As will be discussed below, there are a number of categories which mark past time reference and which are employed only in non-narrative contexts. Sometimes these are extensions either of PFCT or PLPFCT. One salient property of PFCT in general is the fact that it is not used in narrative contexts. This problem will be further discussed in chapter 5.

Below, we shall discuss those categories that are used for marking differences in remoteness of events. As we shall see there, there is a clear tendency for such differences to be neutralized in narrative contexts.

Narrative 'tenses'

As we have already said, some languages employ special TMA categories in narrative contexts (see table 4.7). In our material, clear

Table 4.7 NARR categories

Language	Code	Description	Frequency
Karaboro	D	dependency marker <i>n</i>	27
Kikuyu	ki=	Prefix <i>ki-</i> :suffix 0	17
Sotho	21	Narrative Past	27
Zulu	21	Narrative Past	34

examples are found only in the Niger-Congo phylum: special narrative forms – henceforth NARR – seem to be particularly common in the Bantu family. As is often pointed out in the grammars of those languages, a narrative discourse typically starts with a verb in some non-narrative past form: all the subsequent verbs then obtain NARR. We can exemplify this by the Sotho version of the first connected text of the questionnaire:

(Q.161-5:SO)

Ke ne ke tsamaea ka morung 'I walked (PASTi) in the forest'

Hang-hang ka hata noha 'Suddenly I stepped (NARR) on a snake'

Ea ntoma leoto 'It bit (NARR) me in the leg'

Ka nka lejoe 'me ka bata noha eo 'I took (NARR) a stone and threw (NARR) at the snake'

Ea shoa 'It died (NARR)'

It seems that it is quite common for NARR to exhibit signs of being in some way subordinated or otherwise syntactically connected to the preceding verb. Thus, in Karaboro, the marker of NARR is a particle *n* which is referred to as 'dependency marker'. Cf. the following two sentences:

(Q.56:KB) ù kù '(Is the king still alive? No,) he died'

(Q.165:KB) ké ŋ kù '(in narrative context:) 'It died'

In the Nyuni languages, the subject marker is different in NARR than in the other indicative tenses (cf. the variation between *ke* and *ka*, 'I', in the Sotho examples above). The negative forms are also different, being more similar to the formations used in subordinate contexts. This seems to be the reasons for the treatment of NARR as belonging to the subjunctive mood in some grammars (e.g. for Setswana, Cole 1955).

In Biblical Hebrew, NARR is expressed by a construction involving the conjunction *wa* and – somewhat unexpectedly – the imperfective form of the verb (Johnson 1979).

It is probable that the original function of narrative verb forms was to serve as the second member of constructions of the type 'S and (then) S'. We may note that the Karaboro dependency marker is used also in (Q.137):

(Q.137:KB) Yàlii le tē dye, ú ŋ lētēr syāŋ yrūhū.

'When I came home, he wrote two letters'

Synchronically, however, the sentences in a narrative text should probably be treated as syntactically independent of each other.

The Past

Table 4.8 PAST categories

Language	Code	Description	Frequency		
Polish	2	Past	153		
Hungarian	2	Past	154		
Swedish	2=	Past	134		
Estonian	20	Past	144		
Russian	2=	Past	147		
English	20	Past	102		
Finnish	2=	Simple Past	134		
Hebrew	20	Past Tense	141		
Tamil	2=	Past	84		
Tigrinya	92	Simple Gerund	149		
Amharic	2=	Simple Past	92		
Bengali	20	Past	81		
Czech	20	Past	146		
Greek (Modern)	20	Past	156		
German	20	Past	95		
Fitzr. Cr. Kriol	B	Aux. <i>bin</i> + V	141		
PASTi					
Limouzi	22=	Imperfect	38	.67	.92
Latin	220	Imperfectum	28	.64	.86
Italian	22=	Imperfect	42	.64	.86
Romanian	22	Imperfect	31	.63	.84
French	22=	Imparfait	38	.60	.79
Persian	M2	Imperfect	26	.60	.77
Arabic	K1	Copula + Imperfective	21	.59	.76
Kurdish	22	Past Continuous	16	.59	.75
Amharic	N2	Imperfect	10	.57	.70
Maltese	K	<i>kien</i> + Imperfective	45	.56	.73
Portuguese	22	Impfv Past	46	.55	.72
Georgian	21	Imperfect	38	.53	.68
Sotho	Ne	Aux. <i>ne</i> + Dep. form	37	.53	.68
Alawa	21	Past Continuous	55	.50	.67
Spanish	21	Imperfect	54	.50	.67
Catalan	21	Imperfect	50	.48	.64
Azerbaijani	P	Past Predicative Affix	46	.46	.61
Arabic (Tun.)	K	<i>ka:n</i> + Impfct/Adj.	17	.44	.53
Turkish	Nr	Past Predicative Suffix	21	.43	.52

Table 4.8 continued

<PASTi				
Oromo (Galla)	T	Aux. <i>ture</i>	45	.37 .51
Bandjalang	22	Past indefinite	78	.33 .55
PASTid				
Beja	22	Past	10	
Bulgarian	22	Imperfect	22	
Hindi/Urdu	21	PtP (pred.)	56	
Karaboro	N	<i>naa</i> + V	17	
Punjabi	T20	Inf. + Imperfect of <i>hona</i>	11	
PASTs				
Hindi/Urdu	2=	Past	47	
Bulgarian	20	Aor.=Impfct (for Cop. only)	11	
Karaboro	B	<i>ba</i> + V	48	
Punjabi	20	Imperfect	32	
Afrikaans	2=	Past Copula <i>was</i>	16	
Tigrinya	E	Past Copula	5	
PASTn				
Guarani	20	Past (suffix <i>-kuri-</i>)	117	
Quechua	20	Past (suffix <i>-rqa-</i>)	75	
Wolof	20	Past	71	
Maori	2	<i>i</i> + V	51	
PASTd				
Afrikaans	H1	Aux. <i>het</i> + PtP	145	

PAST is the cross-linguistic TMA category that corresponds to traditional 'past tenses', to the extent that they do not involve any aspectual categories. In fact, PAST is about the only category whose character as a tense is wholly uncontroversial. However, in the majority of the languages in the sample, and probably in the world, past time reference may be indicated in several ways, and in particular by other TMA categories than PAST, most commonly by PFCT, PFV and EXPER. Also, PAST may combine with other categories to form complex TMA categories, such as pluperfects and conditionals. This makes it rather complicated to factor out the exact role of PAST in different TMA systems. For this reason, I have not tried to find in the questionnaire any focal or prototypical uses of PAST or any ranking list of the sort discussed in chapter 2. The assignment of the label PAST to a language-specific category is instead based on checking

'manually' whether the category occurs systematically in those contexts where it should be expected to occur.

The role of PAST, as I hinted above, depends on its interaction with other categories. It appears that PAST is a typical default-choice category in the sense that it is chosen whenever no other past time reference indicating category is eligible. In languages with a PFV:IPFV distinction, PAST categories tend to be used only when a PFV is not used (see chapter 3). In English, the simple past (PAST) is used when the conditions for choosing the present perfect (PFCT) are not fulfilled: here, however, the situation is a bit different since the Past and the Perfect may be combined to yield the Pluperfect. What I have said now is the main reason why an attempt at finding the focal uses of PAST would lead one astray: the distribution of PAST depends largely upon what is left when the other categories have taken their share of the pie.

PAST is predominantly morphologically marked, and in most of these cases suffixally. There are a number of exceptions: in the Semitic languages, PASTi is usually marked by the (past tense) copula. The non-Bantu Niger-Congo languages that have a PAST category tend to mark it periphrastically. In Fitzroy Crossing Kriol, there is a non-bound PAST marker *bin*. In Czech, there is a copula in the first and second persons of the Past, although the third is marked only suffixally. (Polish has traces of this in the enclitic person markers that occur in the same situations.) Still, morphological marking makes up about 75 per cent of the cases of PAST in the sample.

We shall now look at different varieties of PAST that occur in the material. We can say that these types are each characterized by the neutralization of PAST in some contexts.

PASTi – PAST restricted to IPFV

At an earlier stage of the analysis, I assumed that there was a cross-linguistic category that I labelled IMPERFECT (IMPFCT). Later on, I abandoned this hypothesis in favour of the analysis discussed in chapter 3, where the language-specific categories in question are treated as instances of PAST that are restricted to imperfective contexts, and am now using the label PASTi for them instead. A list of these is given in table 4.8. The hypothesized focal uses for IMPFCT are given in tables 4.9-10, although this list will be of less interest under the new analysis.

As we see, this type of PAST is represented by at least 19 languages in the large sample and three in the small sample. Its geographical distribution seems to be restricted to an area around the Mediterranean, including languages from the Afro-Asiatic, Indo-European and Caucasian groups.

Table 4.9 Distribution of PASTi (n>2)

Rank no.	No. of categories	Examples
1	15	101 201 261
4	14	91 111 121 1711
8	13	211
9	12	1611 1661 1811
12	11	281 1161 1181 1402 1761
17	10	981 1022
19	9	1861 1931 1941
22	8	1021 1191
24	7	31 331 1182 1212 1942
29	6	282 321 351
32	5	701 1112 1122 1562
36	4	41 961 1132
39	3	131 661 951 1062 1162

Table 4.10 Prototypical occurrences of PASTi

- Sent. 26 Verb 1 No. of categories: 15
A: Last year, my brother worked at an office. B: What kind of work he DO there?
He WRITE letters
- Sent. 10 Verb 1 No. of categories: 15
A: I went to see my brother yesterday. B: What he DO? (=What activity was he engaged in?)
He WRITE a letter
- Sent. 20 Verb 1 No. of categories: 15
Q: What your brother usually DO after breakfast last summer? A:
He WRITE letters
- Sent. 11 Verb 1 No. of categories: 14
A: I talked to my brother on the phone yesterday. B: What he DO? (=What activity was he engaged in?)
He WRITE letters
- Sent. 12 Verb 1 No. of categories: 14
A: I talked to my brother on the phone yesterday. B: What he DO? (=What activity was he engaged in?)
He WRITE a letter
- Sent. 171 Verb 1 No. of categories: 14
Do you know what happened to my brother yesterday? I saw it myself
We WALK in the forest

Table 4.10 continued

Sent. 9 Verb 1 No. of categories: 14
A: I went to see my brother yesterday. B: What he DO? (=What activity was he engaged in?)
He WRITE letters

PASTs – PAST restricted to stative contexts

This label really covers two kinds of PAST categories. One of these may appear in languages where resultative contexts (see chapter 5) are unmarked. In these languages, the past marker appears in non-dynamic contexts, that is, predications of states or properties that hold at a past time-point or interval. The crucial fact is that this includes contexts where a state in the past is defined by the preceding event that has given rise to it, that is, contexts where a resultative construction could be expected. The latter contexts belong to those where PLPFCTs tend to appear. The only clear example of this type of a PASTs in the material is Karaboro, whose rather intricate TMA system is discussed further in chapter 6. PASTs is more important to the general theory of TMA categories – and maybe to linguistic theory in general – than its rather low frequency in our sample suggests, since it appears to be fairly common among Creole languages, as noted in Bickerton (1981 and elsewhere), who calls this kind of category ‘anterior’. Bickerton connects the anterior with his hypothesis that Creole languages more directly reflect the ‘bioprogram’ that underlies language acquisition than others. In other words, the anterior – or our PASTs – would in some way be more natural than other varieties of PAST. It is rather hard to evaluate a claim of this kind, but the fact that PASTs and similar categories seem to be represented mainly among West African languages, if one is to judge from our material, makes one wonder if the historical connections between these languages and many of the Creole languages discussed in the literature may not be as important here as the hypothesized bioprogram.

The other type of category where the label PASTs has been used are cases where the copula behaves differently with respect to the PAST than lexical verbs.

PASTn

As we have noted above (p. 113), verbs in narrative contexts tend to be unmarked. There are in the sample a few examples of languages where past time reference is marked only in non-narrative contexts. These are Quechua, Guarani, Wolof and Maori. The situation seems

to be clearest in Guarani. In Quechua, there are a number of exceptions to the narrative:non-narrative generalization: PAST is used in the narrative text B3, and unmarked forms are used in quite a few non-narrative examples, predominantly resultative and/or with recent past time reference (e.g. Q.56, Q.141, Q.143 – PAST is used in Q.142 and Q.144). In Wolof, the past marker *won* tends to appear together with the auxiliary *na*, which in isolation is a PFCT (or possibly a RESUL) marker. There are also other examples of original PLPFCTs being used as non-narrative pasts, e.g. in Amharic. The Maori auxiliary *i* is apparently restricted to dynamic non-narrative contexts.

PASTd

There are at least two examples of neutralization of PAST in non-verbal sentences, viz. Beja and Maori (see above). Beja has a construction involving an enclitic copula: this construction is unique in the TMA system of the language by being possible with both present and past time reference. For details, see Dahl 1984b. This label is also used for the cases where lexical verbs behave differently from copular constructions.

Remoteness

Under this heading, I shall treat one of the underlying dimensions of TMA systems, viz. 'temporal distance' or 'remoteness'.³ What I have to say builds less on the TMAQ and more on existing descriptions than the rest of the investigation. Among other things, I draw on the survey of Bantu languages reported in Kopchevskaya 1984.

First, some terminological points. 'Temporal distance' involves, by definition, a measurement of the distance between two points or intervals in time; this implies that for this dimension to be relevant, at least two such time points should be involved in the interpretation of a sentence. Given the Reichenbachian points S, R and E, there are the following possibilities: In the unmarked case, R coincides with either S or E. In those cases, which constitute the overwhelming majority in any text, the only possible distance to measure will be between S and E, that is, 'distant' will mean 'distant from the time of speech'. If R is separate, however, we will have two intervals to measure: on one hand, the distance S–R, on the other, the distance R–E. In principle, both these might be relevant in a TMA system. The tendency, however, seems rather to be for remoteness distinctions to be neutralized in such contexts; many languages do not even have a separate category which like the English Pluperfect is used for events

that take place before an R which in its turn precedes S. I also have relatively little information concerning these cases – being conceptually more complex, they are rather hard to elicit reliable information about – and shall just note one fairly clear example of a minimal pair differing in the distance between a past R and a preceding E. Morolong 1978 quotes the following Sotho sentences saying that the (a) sentence with the *ne-tsoa* form 'is felt to be nearer to the reference point than it is the case with the *ne-stem-ile* sentence in' (b).

(4.7a) Ha letsatsi le-likela re-ne re-tsoa tloha Maseru 'At sunset we had just left Maseru'

(b) Ha letsatsi le-likela re-ne re-tloh-ile Maseru 'At sunset we had left Maseru'

The discussion in the following will mainly concern distinctions connected with the S–E distance.

The E point may be both in the past and in the future: accordingly, languages make distinctions both between 'remote' and 'close' pasts and 'remote' and 'close' future tenses. However, in general, the distinctions in the past appear to be more well developed – that is, to be more numerous and well-defined than those in the future; Ultan 1978 explains this fact by the general tendency for future tenses to be more marked than past tenses. Whatever the facts may be, my material is more extensive on distinctions in the past than in the future, and the bulk of the ensuing discussion will concern the past.

So far, I have mainly used the term 'temporal distance' as a label for the problems that interest me in this section. Since this term is a bit unwieldy, I shall often use 'remoteness distinction' to denote the grammatical categories that are used to mark how far time points are from each other.

Examples of remoteness systems

Remoteness distinctions can be found in languages from most parts of the world and a large number of unrelated genetic groups, although they are more salient in some, such as the Bantu languages. It is possible that temporal distance is at least marginally relevant for TMA categories in the majority of human languages; at present, such a hypothesis is impossible to evaluate due to lack of reliable data. I shall now give a couple of examples of typical remoteness systems, basing the accounts on secondary sources.

A relatively representative example of a well-developed Bantu system of past tenses is found in Kamba (Whiteley & Muli 1962), which distinguishes between three degrees of remoteness in non-narrative contexts:

- (i) an 'immediate' past, 'which refers to an action taking place earlier on the day of speaking': *ningootie* 'I pulled (this morning)';
- (ii) a 'recent' past, 'which refers to an action taking place the day prior to the day of speaking, or even to a week previously': *ninina:kootie* 'I pulled (e.g. yesterday)';
- (iii) a 'far past' tense, 'which occurs for actions having taken place ... not earlier than some month past': *Akamba maia:tua vaa tene* 'The Kamba did not live here in the past'.

In narrative contexts, there is a vaguer distinction between two tenses, Narrative I, e.g. *na:tata*, 'I tried', and Narrative II, e.g. *nina:tata*, 'the same', the latter of which 'connotes a rather less remote time in the past'.

In addition, there is a Perfect, e.g. *ninakoota*, 'I have pulled', 'which may be translated by "has" or "has just"', so that it may have both an immediate – or less commonly – a general perfect connotation'.

As for reference to the future, Whiteley and Muli claim that the same tense, called 'present continuous', e.g. *nunu(k)ukoota*, 'he is pulling, about to pull', is used for referring to the present and to events that will take place within the next 24 hours. There are, in addition, two proper future tenses, one simply called 'future', e.g. *aka:koota*, 'he will pull', which is 'for events occurring subsequently to the time of speaking up to a period of some months ago' (41) (it is not clear if this is supposed to exclude 'today': in another place the future tense is defined as 'from 24 hours beyond the time of speaking' (49)), and a 'far future', e.g. *nitukaatata*, 'we shall try', 'used for actions taking place at some point after a few months, though it is clear that there is some looseness in this'.

For an example of a system from another genetic group and another part of the world we may quote Derbyshire's description of Hixkaryana, a Carib language (1979). According to Derbyshire, Hixkaryana has – like Kamba – three degrees of remoteness in the past: 'immediate past', e.g. *kahatakano*, 'I came out', which 'refers to actions done the same day or the previous night', 'recent past', e.g. *ninikyako*, 'he went to sleep', which 'refers to actions done on the previous day or any time earlier up to a period of a few months (this is the norm, but it is relative to the total situation, and sometimes an event of only a few weeks ago will be expressed with the distant past suffix)', and 'distant past', e.g. *wamaye*, 'I told it', which 'refers to actions done any time earlier'. (Hixkaryana has no future tenses.)

As we can see, the similarity between the ways in which the two languages 'cut up' the past is striking.

A much more complicated system is found in Kiksht (Hymes 1975). Two sets of morphemes, one with four members and one with two,

interact to yield at least seven (possibly more) past tenses. In addition to the distinction 'today: before today', Kiksht has a distinction 'this year: before this year' ('seasonal round' in Hymes's terminology) and a possibility in each 'slot' to have a finer gradation between 'near' and 'far'. Thus, in the 'before this year' range one may, according to Hymes, distinguish between what is within the realm of personal experience and what belongs to the 'age of myth'. Within 'this year' one can distinguish 'not more than a week ago' (or some equivalent time measure) from 'more than a week ago'. In the future, there is simply a distinction between 'near' and 'far', although no indication of what that means is given. Until evidence to the contrary appears, it seems fairly safe to regard Kiksht as an example of maximal differentiation in remoteness systems, at least with regard to the past.

Objective and subjective judgements of temporal distance

When someone assesses the distance of an event from the present time-point, his/her judgement may of course be influenced by a number of more or less conscious factors. We can range judgements of temporal distance on a 'subjectivity scale', where the zero point is a measurement which is made purely in physically definable terms. Most (if not all) cultures employ some kind of physically defined time measures, usually based on the observable reflections of the movements of the earth, the sun and the moon. *Day*, *month* and *year* are all definable as 'one cycle' in different cyclical astronomical processes. Terms like *hour*, *minute* and *second*, which are defined as subdivisions of days depend on the existence of reasonably reliable instruments for the measurement of time. Adverbial expressions which denote time-points or intervals defined in terms of these measures are of course extremely common in English and probably most other languages, e.g. *yesterday*, *four months ago*. A sentence such as the following would then exemplify an objective time measure:

(4.8) I arrived here exactly two years ago

As a contrast, the time referred to in the following sentence could – depending on the circumstances and the mood of the speaker – vary between, say, ten minutes and sixty years:

(4.9) I've been here for an awfully long time already!

As we have already seen, objective time measures do play an important role in determining the choice between different tenses in various languages. However, it appears that there are differences between languages as to how important they are. In general, there seems to be some possibility for the speaker to treat something as close even if it is objectively remote and vice versa, that is there is a possibility to give weight to subjective factors. In some languages,

however, a 'contradictory' combination of, say, a 'today' tense with a time adverbial meaning 'last year' results in an ungrammatical sentence. This appears to be the case for instance in Kom (a Grasslands Bantu language, Chia 1976). In other languages, e.g. Sotho, it seems in general possible to combine any time adverbials with any tense (Morolong 1978), as in the following sentence where a recent past is used:

- (4.10) Morena Moshoeshoe ofalletse Thaba Bosiu ka-1824
'Chief Moshoeshoe moved to Thaba-Bosiu in 1824'

In other words, it appears to be possible to distinguish between those languages that give more weighting to objective factors and those which leave more room for subjective factors in judgements of temporal distance. Such factors may be for instance spatial distance or personal involvement. Consider the following two Limouzi sentences from Javanaud 1979:

- (4.11) I m'an letsa quant j'ai paia quo qu'i devio
(4.12) I me latseren quant i'agui paia quo qu'i devio 'They released me when I had paid what I owed'

The verbs in the main clauses of (4.11–12) are in the *passé composé* and the *passé simple*, respectively. According to Javanaud, (4.11) is appropriate if 'we are still at the same place'. Similar intuitions were elicited for the distinction between hodiernal and hesternal pasts in Sotho. Intuitively, it is not too hard to accept that distance in time and space will not always be differentiated in people's minds. In a parallel way, events which you have witnessed yourself or which concern you as a person in a direct way might be felt as being 'closer' in a general way and thus be more likely to be reported in a non-remote past tense. (Notice that spatial distance may be measured either relative to the point where the speech act takes place or relative to where the speaker or some other protagonist of the conversation was situated when the event took place. In the second case, there is a clear connection between what is spatially close and what is witnessed by the speaker.)

Colarusso 1979 discusses the use of certain prefixes in some North-West Caucasian languages, which he takes to mark what he calls 'horizon of interest'. The main verb of a sentence is marked for 'horizon' when 'the action referred to takes place or originates' in a zone 'lying at a variable, culturally determined distance' from some central locus. This may be interpreted either in concrete spatial terms, in such a way that 'marked for horizon' means 'spatially distant from the speech act', or in abstract terms, in which case 'horizon' refers to, for instance, what directly concerns the speaker. More specifically, it may indicate that one is speaking of a person who is not within the speaker's primary social group, i.e. his consanguineal kingroup

(hence the reference to 'kinship' in the title of the paper). The relevance of Colarusso's paper for the present discussion is that it shows how a category which primarily marks one type of distance may be extended to other types. The prefixes discussed by Colarusso do not have temporal uses, as far as one can judge from the data given in the paper. An example of a language where morphemes with a deictic spatial meaning have acquired temporal meaning is Kiksht as described by Hymes, where the prefixes *t-* and *u-* have the primary meaning 'hither' and 'thither' and the secondary use of marking the time of reference as being relatively closer or more distant, respectively.⁴

The hodiernal:non-hodiernal distinction

There is evidence for the following generalization:

- (4.13) If there is one or more distinctions of remoteness in a TMA system, and reference can be made to objective time measures, one of the distinctions will be between 'more than one day away' and 'not more than one day away'.

The distinction between 'today' and 'before/after today' tenses, which we shall refer to as the hodiernal:non-hodiernal distinction (from Latin *hodie*, 'today'), is well known from grammars of Bantu languages, but is by no means restricted to this group: we have already seen illustrations from Hixkaryana and Kiksht, and we may refer to Davies's (1981) description of Kobon, a New Guinea language, to get an example from another geographical area not represented in our sample. Turning to the Romance group within the Indo-European languages, we find in the Port-Royal grammar (Lancelot & Arnauld 1660; 108–9) a clear statement to the effect that the categories *passé composé* and *passé simple* differed in that the former was used as a hodiernal past and the latter as a pre-hodiernal, and even if this distinction has now been lost in Modern French, it is still reflected in Spanish, Catalan and Occitan. In the TMA questionnaire, the exchange of the word *yesterday* for *this morning* in sentences (141–4) resulted in a different choice of tenses in the following languages, as well as the ones we just mentioned: Bengali, Kikuyu, Quechua, Zulu.

Sometimes, grammars give a more specific delimitation of what is counted as 'today', e.g. when it is said in a grammar of Ewondo (a Bantu language) that 'aujourd'hui commence au dernier coucher du soleil' (Angenot 1971). When a day is supposed to begin is clearly a culture-bound phenomenon, and statements like the one quoted may be regarded as language-specific sharpenings of a universal but imprecise boundary (to the extent that they are not just constructions by the grammarian, of course).

Typically, then, there are one or more hodiernal tenses and one or more pre-hodiernal ones. As to delimitations among the pre-hodiernal tenses, they are most often much vaguer. A typical description is the one quoted above for Kamba; referring to 'the day prior to the day of speaking, or even to a week previously'. It appears that if there are two pre-hodiernal tenses in a language, the marked member of the opposition is a 'distant past' which typically refers to things that happened several months or years ago. The unmarked member would then be the tense often referred to as the 'yesterday tense', although, as we have already seen, it tends to go further back than yesterday. For such forms, I have coined the label 'hesternal' (from the Latin *hesternus* 'related to yesterday').

Since the delimitation between hesternal and distant pasts is usually rather imprecise, one might question whether it is at all relevant to think of it in terms of objective time measures. It would then be tempting to strengthen (4.13) to state that the only relevant objective time measure is that of 'one day'. However, Hymes's account of Kiksht, related above, suggests that at least for some languages, the year may be another relevant measure. Interestingly, even in Kiksht, with its rich remoteness system, something like 'hodiernal:pre-hodiernal' may well be the fundamental distinction. Hymes notes that the hodiernal past *i(g)*- 'appears to be far and away the preferred tense for recent past and to be used as such in conversation and narrative' and that it as such 'is contrasted with *ga(l)*- as the preferred tense for distant past'. 'One gains the impression that the first "cut", so to speak, made by speakers in terms of times past is *recent* (*i(g)*-): *remote* (*ga(l)*-)'.

(4.13) seems to hold also for remoteness distinctions in the future, although it is relatively more seldom that objective time measures are found to be relevant there at all. As an example of a language which does have an opposition between hodiernal and post-hodiernal future, and where the hodiernal future is distinct from the present (in contrast to e.g. Kamba), we may mention Aghem (Bantu, Hyman 1979).

Remoteness distinctions and narrativity

As we suggested above, marking of temporal distance may be different in narrative and non-narrative contexts. We have already seen that Kamba has three distance distinctions among its non-narrative tenses, whereas it has only two in the narrative ones. Furthermore, it was noted that the distinction between the two narrative tenses seems much vaguer. Another example would be Limouzi. Here, as in the surrounding Romance dialects and in older forms of French, *passé composé* tends to be used as a hodiernal past

and *passé simple* as a pre-hodiernal past. This holds only for non-narrative contexts, however: in narrative contexts, only *passé simple* is possible. This suggests the following generalization:

- (4.14) If narrative and non-narrative contexts differ with respect to the marking of temporal distance, it will be the non-narrative contexts that exhibit the largest number of distinctions.

One explanation for (4.14) might lie in the character of time reference in narrative contexts: since the point of reference is by definition determined by the context, any further indication of its location in time will be redundant. Another explanation might be valid for those languages where PFCT and hodiernal past are non-distinct (a problem that we shall return to in chapter 5). The propensity for hodiernal pasts to be used in non-narrative contexts could be a consequence of the non-narrative character of PFCT, which will be discussed in chapter 5.

Immediate past tenses

In many languages, including those that do not otherwise mark remoteness distinctions systematically, there are constructions that may be used to translate the English Perfect with the adverb *just*, as in *He has just arrived*. In the Romance languages we thus find constructions like the French *venir de* + infinitive (literally 'to come from doing sth.'). The semantics of these constructions is not quite clear; although it might be tempting to assume that they express a stronger closeness than a hodiernal past, it rather appears that the 'immediacy' involved is generally not measurable in objective terms, which could be taken to mean that these constructions are, strictly speaking, outside the system of more objective remoteness distinctions. Consider e.g. a sentence like *The age of computers has just begun*.

In our material, 'immediate pasts', as the constructions are called, by and large play a marginal role, with the exception of Spanish and Catalan, where the construction *acabar de* is used sufficiently often to qualify as a 'major TMA category'.

Remoteness categories in the sample

In most cases, remoteness has been regarded as a secondary feature of some TMA category, mostly PFCT or PLPFCT (for discussion, see p. 136). Except for the immediate past constructions in Spanish and Catalan, remoteness categories have only been used as labels for 'major TMA categories' in the three Bantu languages Kikuyu, Sotho

and Bantu. The material at hand is not sufficient for a reliable labelling of the categories in question: no enumeration is therefore given here, but the reader is referred to table 6.22 (p. 175). Perhaps the most important remaining problem is the relation between remoteness and the category of PFV:IPFV.

5

The Perfect (PFCT) and its relatives

The semantics of the categories labelled 'Perfect' in e.g. English and Swedish belong to the most discussed problems in the theory of grammatical categories. Claims as to the existence of a universal category PERFECT have been met with scepticism based on the limited empirical data available and on the difficulties in finding a definition of the category. I regard it as one of the major results of this investigation that it is now possible to postulate with some confidence such a cross-linguistic category of PERFECT (PFCT).

In this chapter, we also treat some other categories which are closely related to PFCT, viz. the EXPERIENTIAL (EXPER) and the PLUPERFECT (PLPFCT), and also the category QUOT, which – being primarily a modal category – does not fit in very well anywhere else.

The Perfect (PFCT)

PERFECT (PFCT), with its subvarieties, occurs in a wide range of languages. Clear examples (as listed in table 5.1) are found in at least 24 languages in the material and five languages (27 per cent) in the small sample. Categories whose extensions are included in PFCT and which are not identified as any other category occur in another nine languages in the material and two languages in the small sample (11 per cent).

PFCT is rather consistently marked periphrastically (about 85 per cent of the cases), the only clear counterexamples being Niger-Congo languages (Akan, Kikuyu). Typically, constructions involving a copula or some auxiliary together with some past participle or similar form of the verb are used. For some languages, a particle analysis may be more natural (West African and Malayo-Polynesian groups) but the criteria are not always clear. In Yoruba, the particle used in PFCT contexts also has the (basic?) interpretation *already* – there may well be more cases of a PFCT derived from that kind of source. In Karaboro, the particle *yaa*, 'already', occurs in a number of cases

Table 5.1 PFCT categories

Language	Code	Description	Frequency	C	HR
PFCT					
Swedish	H	Aux. <i>ha</i> + Supine	61	.64	.87
English	H	<i>have</i> + PtP	37	.61	.81
Kammu	H	<i>hóoc</i> + V	24	.61	.79
Limouzi	H	<i>o</i> (<i>eitre</i>) + 92	54	.61	.81
German	H	Aux. <i>haben/sein</i> + PtP	64	.59	.80
Maori	pf	<i>kua</i> + V	27	.58	.74
Indonesian	S=	<i>sudah</i> + V	21	.57	.00
Spanish	H	Aux. <i>haber/estar</i> + PtP	55	.56	.75
Isekiri	R	V + <i>re</i>	18	.54	.67
Catalan	H	Aux. <i>haver/esser</i> + PtP	56	.53	.71
Javanese	W	Aux. (<i>u</i>) <i>wis</i> + V	24	.53	.67
Akan	p=	Perfect	22	.52	.64
Finnish	O	Aux. <i>olla</i> + Sup.	62	.51	.69
Punjabi	A	PtP + Aux. <i>hona</i>	51	.50	.67
Thai (Siamese)	L	V + Aux. <i>lææo</i>	26	.50	.62
Sundanese	Pa	<i>parantos</i> + V	23	.49	.61
Estonian	O	Aux. <i>ola</i> + PtP	49	.49	.65
Bulgarian	24	Copula + Aor. PtP	35	.49	.63
Bugis Makassar	L	<i>leba</i> + V	28	.49	.61
Yoruba	p=	<i>ti</i> + V	26	.47	.58
Wolof	Na=	V + <i>na</i>	29	.44	.55
Amharic	H1	Perfect	19	.44	.53
Tamil	I0	PtP + Aux. <i>irykka</i>	24	.41	.50
<PFCT					
Greek (Modern)	E	Aux. <i>écho</i> + Supine	21	.39	.48
Hindi/Urdu	A	PtP + Aux. <i>hona</i>	63	.38	.56
Kikuyu	it=	Suffix <i>-it-</i>	16	.37	.44
Bengali	A	PtP + (contracted) copula	80	.35	.58
Greenl. Eskimo	2	Suffix <i>sima</i>	27	.32	.41
Oromo (Galla)	er	Passato prossimo	8	.22	.25
Czech	M	<i>mít</i> + Pass. PtP	6	.14	.17
Beja	C	PtP + enclitic copula	12	.12	.17
Portuguese	T	<i>ter</i> + PtP	12	.12	.17
PFCTq					
Persian	Bu	PtP + Aux. <i>budan</i>	50	.62	.00
Azerbaijani	22	Suffix <i>-mIsh</i>	68	.55	.75
Georgian	p	Perfect	50	.48	.00

where there is no counterpart to it in English, but this does not happen often enough to warrant an analysis of *yaa* as a PFCT.

The most important variety of PFCT is the PERFECT-

QUOTATIVE (PFCTq), i.e. a category which differs from the ordinary PFCT by being used also as a quotative. It appears in Azerbaijani, Georgian and Persian in our material and is discussed in more detail below.

Semantics of PFCT

Table 5.2 Distribution of PFCT ($n > 3$)

Rank no.	No. of categories	Examples
1	31	531
2	28	541 561
4	27	421 641
6	26	671
7	25	371 1341 1361 1392
11	24	1331 1351 1531
14	23	431 591 1382
17	22	481 891
19	21	511 521 571 681 1491 1501
25	20	551 901
27	19	411 1291
29	18	1072
30	17	691 1082
32	16	441 491 631 1541
36	15	1271 1281 1301
39	14	401 611 1511
42	13	501 581 601 621 1051
47	12	1052 1061
49	11	1431 1481
51	10	391 1062
53	9	381 651 1132 1411
57	8	451 1441
59	7	461 471 661 701 851 1421
65	6	1112 1771 1781

Table 5.3 Prototypical occurrences of PFCT

Sent. 53 Verb 1 No. of categories: 31

(A: I want to give your brother a book to read, but I don't know which. Is there any of these books that he READ already?)

B: (Yes,) he READ this book

Sent. 54 Verb 1 No. of categories: 28

A: It seems that your brother never finishes books.

(That is not quite true.) He READ this book (=all of it)

Table 5.3 continued

Sent. 56 Verb 1 No. of categories: 28

Q: Is the king still alive? A:
(No,) he DIE

Sent. 42 Verb 1 No. of categories: 27

Q:
You MEET my brother (at any time in your life until now)?

Sent. 64 Verb 1 No. of categories: 27

Child: Can I go now? Mother:
You BRUSH your teeth?

Sent. 67 Verb 1 No. of categories: 26

Q: What did you find out when you came to town yesterday? A:
The king DIE

Sent. 136 Verb 1 No. of categories: 25

A person who has heard (135) but not seen the event says:
The king ARRIVE

Sent. 139 Verb 2 No. of categories: 25

When I COME home (yesterday), he WRITE two letters (=that is what he accomplished during my absence)

Sent. 134 Verb 1 No. of categories: 25

A person who has heard (133) but not seen the event says:
The king ARRIVE

The distribution and prototypical uses of PFCT are listed in tables 5.2-3. Commonly (McCawley 1971, Comrie 1976, 56), the English Perfect is said to have the following uses:

- (i) perfect of result (also called 'stative perfect') – typical examples: (Q.54), (Q.69). See table 5.3.
- (ii) experiential (or 'existential') perfect – typical example: (Q.42). See table 5.3.
- (iii) perfect of persistent situation – typical example: (Q.148)
Context: (Of a coughing child:) For how long has your son been coughing?
Sentence: He COUGH for an hour
- (iv) perfect of recent past (also called 'hot news' perfect) – typical example:

(Q.133)

Context: The speaker has just seen the king arrive (an unexpected event)

Sentence: The king ARRIVE

We shall not discuss here the issue of whether the English Perfect exhibits ambiguity in the proper sense of that word (treated in e.g. McCoard 1976) – it is not clear if this is a meaningful question for TMA categories. Since (i–iv) in any case represent distinguishable types of situations in which grammatical categories such as the English Perfect may or may not be used, we shall include them in our conceptual repertoire, although with some reservations, which follow.

Comrie refers to (i–iv) as 'types of perfect'. This way of speaking reflects the two-level model he uses, where cross-linguistic semantic categories and cross-linguistic grammatical or functional categories are not distinguished: 'types of perfect' sounds as if we were dealing with different kinds of grammatical categories, although what they primarily are is types of *uses* of such categories. In fact, it turns out that categories that appear to exhibit only one of the typical uses of PFCT are often subtly different from PFCT in their semantics, as we shall see below.

One further reservation about (i–iv) regarded as types of situations in which PFCT may be used is that these sets are not necessarily disjoint. In particular, (i) and (iv) overlap quite considerably. As I will argue below, this is in fact what makes it possible for these uses to be found in one category.

Since the 'readings' of PFCT co-occur so often in one and the same language-specific category, the question naturally arises what unites them. It seems to me that the common factor – which will be as close to a 'Gesamtbedeutung' for PFCT as we can come – is that both involve a point of reference (in Reichenbach's sense) which is different from the 'point of event', although the role of the R will be slightly different in each case.

The 'perfect of result' A typical characterization of the perfect of result is given by Comrie (1976, 56): 'a present state is referred to as being the result of some past situation'. This formulation hides an equivocation of some importance. To see this we have to look at constructions other than those traditionally labelled 'perfect' that also seem to convey a meaning characterizable in the same way.

Nedjalkov et al. (1983) distinguish three types of categories: (i) statives, which are used to denote states simpliciter; (ii) resultatives, which are used of states which presuppose a preceding event; (iii) perfects, i.e. what I call PFCT. They point out that resultatives differ

from perfects in the ways in which they can be combined with temporal qualifiers. I shall illustrate that by an example from Swedish. In addition to the PFCT formed with the auxiliary *ha* + the so-called supine form of the verb, there is in Swedish a construction with *vara*, 'be', followed by a past participle. This construction is 'ergative' in the sense that the subject is interpreted as the 'deep' subject for intransitives and as the 'deep' object for transitives. Thus, (5.1) means that someone has gone away and has not returned yet.

(5.1) Han är bortrest (lit.) He is away-gone'

The question is now: Is (5.1) synonymous to (5.2) in its resultative use?

(5.2) Han har rest bort 'He has gone away'

If we accept the characterization of this use quoted above, there is no reason why it should not be. However, if we add the adverb *fortfarande* to (5.1) and (5.2), we see that there is a difference in the resulting sentences: (5.3) is quite normal, but (5.4) is hardly acceptable.

(5.3) Han är fortfarande bortrest 'He is still away-gone'

(5.4) ??Han har fortfarande rest bort 'He has still gone away'

(If (5.4) can be used at all, it would according to my intuitions mean something like 'I am tired of your asking for X; I am telling you that he has not returned yet.')

How can we explain this difference? It appears to me that the perfect used in a resultative sense differs from a resultative construction like (5.3) in that there is more focus on the event than on the state. This intuition is rather hard to capture in formal terms: truth-conditional semantics of the standard brand has no place for things like 'emphasis' or 'focus'. One might however venture an account of the following kind. Whereas the stative construction in (5.1) might well be described in terms of a 'present state resulting from an earlier event', a statement containing a 'perfect of result' should rather be characterized as being made against a background state of affairs in which the event referred to in the sentence has not yet taken place. What is said then would be that the present state of affairs differs from the background one by the event's taking place. Words like *fortfarande*, 'still', on the other hand, also presuppose a background state of affairs but are used precisely to indicate the lack of a difference between that state-of-affairs and the actual one. This contradiction between the semantics of the 'perfect of result' and that of words like *still* would explain why they do not occur together.

The term 'result' may be understood in a wider or narrower sense. In the wider sense, everything that is caused by an event may be said to be the result of that event. For instance, my happiness may be a

result of my getting a pay rise. In the narrower sense, a state is said to be the result of an event if that state is part of the characterization of a certain event-type to which the event is said to belong. For instance, if a person dies, the result in the narrow sense is that he is dead: the results in the wider sense include e.g. his relatives being sad. It is results in the narrower sense that characterize resultative constructions: in the case of PFCT, the delineation is much harder to draw, and that is one of the reasons why the resultative use of the perfect shades off into the experiential use. It follows from what I said in the preceding paragraph that resultative constructions can only be formed from verbs whose semantics involves change of some kind. In general, resultative constructions tend to be highly lexically restricted. There also seems to be a high correlation between passive voice and resultative constructions: indeed, in many languages, resultative constructions are only found in the passive voice or some functional equivalent of it – an example would be the Russian construction with a past passive participle in predicate position exemplified in the following sentence:

(5.5) Dom postroen 'The house is/has been built'

Often, as indeed in Russian, such constructions are ambiguous between a resultative meaning and a 'dynamic' interpretation. The same holds true of the English passive construction (*be* + past participle), although the resultative use is at least synchronically rather marginal.

It is rather hard to identify resultative constructions in our material: one reason is that passive constructions are not well covered in the questionnaire for reasons discussed in chapter 2. Also, there are hardly any sentences that would be criterial for distinguishing resultatives from PFCT or stative constructions. The book by Nedjalkov et al. covers 'resultative constructions' – in the wider sense which includes statives and perfects – in more than twenty languages: it turns out, however, that among those there are very few examples of 'active' resultatives which are distinct from both perfects and statives.

One possible candidate for a RESUL category is the Japanese *-te iru* construction (in one of its uses, see chapter 6). Sentence-final *le* in Chinese might also be a RESUL, although it has been interpreted as a PFCT (Li et al. 1982).

Some categories are used in both stative and PFCT contexts: it appears that this is not too uncommon for default categories such as, in our material, the unmarked categories in Karaboro and Quechua. Another case is the so-called 'perfective aspect' in Oneida. One might expect that these categories would not be used in 'experiential' contexts, but they clearly are in our material. Thus, they should probably be analysed as PFCT/STAT rather than STAT/RESUL.

PFCT and recency The statement that PFCT and HOD-PAST are often non-distinct expresses a fairly well-known tendency. The following quotation represents a more traditional way of expressing the same thing:

In many languages, the perfect may be used where the present relevance of the past situation referred to is simply one of temporal closeness, i.e. the past situation is very recent. (Comrie 1976, 60)

Welmers (1972, 348) goes as far as to claim that many so-called recent or hodiernal pasts in African languages are really what he calls 'completive', i.e. 'perfects of result'. In making this claim, Welmers does not consider the possibility that temporal closeness and having a result at the point of speech may well both be relevant factors for one and the same category, something which is at least fairly clear for some of the languages in our material, such as Limouzi. It should be pointed out here that the identification of PFCT and HOD-PAST is by no means an absolute universal: exceptions among the Bantu languages are Kikuyu, apparently Kamba (see above) and LoNkundo, which Welmers himself cites, in all of which there are separate categories. The non-distinctness of PFCT and HOD-PAST is, however, frequent enough to merit an explanation. An obvious one would be that a recent event is more likely to have a persistent result than a distant one. A category which is used in the sense of a 'perfect of result' will thus automatically be used more frequently of recent events.

It may be somewhat difficult to distinguish such a tendency from another possible restriction on the use of perfects, namely that if they are used to state that a certain kind of event has taken place within a certain temporal frame, that interval must not be ended before the point of speech. This restriction has been the basis for the 'extended now' theory of PFCT (McCoard 1976.) In accounts of remoteness categories in Bantu grammars, reference is sometimes made to 'current' and 'preceding units of time', where the units of time may be days, months, years or even wars (Appleby 1961, quoted in Johnson 1981). Thus, a hodiernal past would refer to 'the current unit of time', variously interpreted as 'today', 'this week', 'this year' etc., whereas a hesternal past would refer to 'yesterday', 'last week' etc. The day as a unit of time would then presumably represent a default value.

The 'universal perfect' or 'perfect of persistent situation' The interpretation of PFCT exemplified by (Q.148) has been referred to variously as the 'universal perfect' (McCawley 1971) or 'the perfect of persistent situation' (Comrie 1976; 60):

(Q.148:EN) He has been coughing for an hour

In many languages, a present tense or an unmarked category is used in these contexts, cf. e.g. Sundanese:

(Q.148:SD) Anjeunna batuk sajam

There are, however, quite a few examples in the questionnaire of PFCT here. The list includes the following languages: Amharic, English, Estonian, Finnish, Kammu, Kikuyu, Limouzi (present tense also given as alternative), Spanish, Swedish and Wolof.

PFCT, definiteness and narrativity The English Present Perfect does not in general go very well together with definite time adverbials: a well-known fact that is illustrated by the unacceptability of a sentence such as (5.6) (cf. (Q.141-4)):

(5.6) *I have met your brother yesterday

We noted above that there is at least one clear exception to this, viz. when the temporal adverb refers to a current temporal frame, e.g. *today*. As various authors have pointed out, the constraint does not just concern explicit time adverbials but definite time reference in general. Thus, (5.8) would be unacceptable as an answer to the question (5.7).

(5.7) What happened at two o'clock yesterday afternoon?

(5.8) I have met your brother

Although some scholars have taken the restriction to indefinite time reference to be the essential characteristic of the Perfect in English, it is not possible to make it a universal restriction on PFCT. It is true that PFCT has a low rate of occurrence in sentences like (Q.141-4) – especially in (Q.142), where proximity in time and space does not help. However, some languages clearly do differ from English in this respect. For instance, Bulgarian grammars quote sentences like the following as perfectly grammatical:

(5.9) Pratkata e pristignala predi tri dena 'The delivery arrived (PFCT) three days ago'

In Swedish, although the literal translation of (5.6), (5.10), is about as unacceptable as its English counterpart, the restriction turns out to be much less water-tight than one might think at first blush.

(5.10) Jag har mött din bror igår

Sentences containing definite time adverbial and verbs in the Present Perfect do occur from time to time in both written and spoken Swedish, and they cannot always be interpreted as instances of an 'afterthought' construction, as is sometimes proposed as an explanation of the fact that (5.6) is possible in English if you have an intonation break before *yesterday*. The information structure of the

sentence seems to be relevant, however. Jespersen (1924) claims that a definite time adverbial can co-occur with the Danish Perfect if it is not in sentence-initial position.¹ At least for Swedish, I think that the relevant factor is not so much position as whether the time adverbial represents 'new information' in the sense that it refers to a point or interval in time which has not been considered earlier in the context. Such a time adverbial will normally bear full stress and is more likely to appear in non-initial position. Thus, a Swedish sentence like (5.11) with the main sentence stress on the time adverbial will not be experienced as deviant to any significant degree.

(5.11) Jag har besökt England i januari nittonhundrafyrtiotvå
'(lit) I have visited England in January, nineteen forty-two'

Formulating the restriction in terms of information structure is in my opinion consistent with a Reichenbachian theory of PFCT: the crucial thing in that theory is whether there is an R which is separate from the E. The need for such a separate point of reference will occur only if the E is not already 'given' in the context. It is possible that the restriction can vary in how far it goes: apparently, English belongs to the stricter languages here. In fact, there may be a hierarchy of definiteness in temporal reference, which is such that when you ascend it the probability of using a PFCT gradually diminishes. The high end of it would then be represented by narrative contexts, as defined in chapter 4. Looking at the occurrences of PFCT categories in narrative contexts in the questionnaire, we see that they are very low indeed (with the exception of quotative contexts). We may interpret this as due to the self-sufficiency of narrative contexts with regard to time reference: since time is (by definition) established by the immediately preceding linguistic context, the need for any external reference-points is nil. (For clarity, it should be added that narrative texts may of course contain perfects – perhaps more typically pluperfects – having the function of creating 'flashbacks' to a time earlier than that defined by the narrative time, which then takes over the role of R.)

PERFECT vs. PERFECTIVE

In linguistic literature, the terms 'perfect' and 'perfective' are often used interchangeably, and even if this usage is not so often found in serious treatments of tense and aspect, it happens quite often that otherwise well-oriented linguists are astonished to find that there may be a difference. As we saw in chapter 2, the two cross-linguistic categories PERFECT and PERFECTIVE are so different in their distribution that the possibility of confusing them in our material is very small. Still, in view of the prevailing uncertainty about the status

of the two categories, and the fact that there is admittedly a considerable overlap in their distribution, it may be wise to say a few words about how you can tell one from the other.

If one looks at the distribution in the questionnaire, the most striking difference is in narrative contexts: as we have noted above, PFCT is very rarely used narratively, whereas several of the prototypical cases of PFV belong to these contexts. This can be generalized: the reluctance we have seen against using PFCT with definite time reference is completely absent in PFV.

We may also note that PFCT can in many languages be used in combination with progressive constructions, as in *I have been sleeping*, something not to be expected if PFCT expressed perfectivity.

If we look at the ways in which PFCT and PFV are marked, we can note that PFCT is overwhelmingly marked syntactically, whereas PFV is in the majority of cases marked morphologically.

In spite of what we have said about the differences, however, it should be pointed out that historically, it is not uncommon for a PFCT to develop into a PFV (the other direction seems less well attested). This has happened e.g. in a number of the Romance languages, such as French. The nature of this process is not clear: it may well be that it can be subsumed under a general tendency for peripheral TMA categories to be attracted towards the centre of the TMA system, acquiring in the process properties that are characteristic of central categories.

The EXPERIENTIAL (EXPER)

Table 5.4 EXPER categories

Language	Code	Description	Frequency	C	HR
EXPER					
Japanese	Kg	<i>koto ga aru</i> construction	8	.66	.88
Javanese	T	<i>tahu</i> + V	8	.66	.88
Thai (Siamese)	Kh	Aux. <i>kheey</i> + V	7	.65	.86
Chinese (Man.)	G	Particle <i>guo</i>	23	.63	.83
Sundanese	K=	<i>kantos</i> + V	10	.62	.80
Indonesian	P	<i>pernah</i> + V	12	.59	.75
Işekiri	Ri	V + <i>re</i>	5	.51	.60
Sotho	Ka	Aux. <i>ka</i> +Dependent form	9	.47	.56
<EXPER					
Cebuano	u=po	Prefix <i>maka-</i>	11	.39	.00

Table 5.5 Distribution of EXPER

Rank no.	No. of categories	Examples
1	6	431
2	5	391 411 421 441 511 521
8	4	481 501
10	2	491

Table 5.6 Prototypical occurrences of EXPER

Sent. 43 Verb 1 No. of categories: 6

Q: Did you know my father, who died last year?
(Yes,) I MEET him (at least once)

Sent. 51 Verb 1 No. of categories: 5

Q: When you came to this place a year ago, did you know my brother?
(No,) I not MEET him (before I came here)

Sent. 39 Verb 1 No. of categories: 5

Q: Do you know my brother?
(Yes,) I MEET him (once) several years ago

Sent. 41 Verb 1 No. of categories: 5

Q: Do you know my brother?
(No) I not MEET him (in my life)

Sent. 44 Verb 1 No. of categories: 5

Q: Did you know my father, who died last year?
(Yes,) I MEET him (several times, now and then)

Sent. 52 Verb 1 No. of categories: 5

Q: When you came to this place a year ago, did you know my brother?
(No,) I not MEET him (before I came here but I met him later)

Sent. 42 Verb 1 No. of categories: 5

Q: You MEET my brother (at any time in your life until now)?

In the material, EXPER occurs in eight languages (see table 5.4) from several different areal and genetic groups, although with a concentration in Africa and East Asia. The examples quoted in the literature are usually either the same as the ones on our list or else from the same geographical areas. In other words, even if it is not possible to regard EXPER as an areal phenomenon, its occurrence seems to be partly conditioned by areal influence.

The distribution and prototypical uses of EXPER are listed in tables 5.5–6. The average frequency of the EXPER categories in the languages listed is 10.4, which makes it a relatively uncommon category in the material. It is noteworthy that all the examples in the material are of periphrastic constructions – there is no language in the sample which uses a morphological process to mark EXPER. These facts clearly mark off EXPER as a relatively peripheral category in the TMA systems in which it appears.

EXPER is a very homogeneous category with one exception – the Chinese *guo* construction. As can be seen from the table, its frequency in the questionnaire is about twice that of EXPER in any other language. In fact, the lower part of the list in table 5.6 is entirely due to Chinese – its high correlation coefficient may be partly an artefact dependent on this circumstance. Actually, there is good reason to assume that Chinese *guo* represents a distinct subvariety of EXPER, if not a distinct category altogether.

Semantics of EXPER

The basic use of EXPER is in sentences in which it is asserted (questioned, denied) that an event of a certain type took place at least once during a certain period up to a certain point in time.

The fullest description to my knowledge of the semantics of an EXPER category in an individual language is found in Inoue 1975, where the behaviour of the Japanese *-ta koto ga aru* construction is described at some length. Since the mutual correlations between the categories in table 5.4 are quite high, it is reasonable to assume that most of what she says will be valid also for the other languages where EXPER is found. Inoue gives the following examples from Japanese:

(5.12) Mike wa, Nihon de hataraita koto ga aru 'Mike has worked in Japan'

(5.13) Kono hashi wa, kowareta koto ga aru 'This bridge has been destroyed'

(5.14) John wa, futotte ita koto ga aru 'John has been fat'

In all these sentences, she says, 'it is implied that the activity, event or state obtained at one or more points in a stretch of time extending from the past to the present' and moreover, 'that the activity, event or state is repeatable'. The latter can be reinterpreted as saying that the sentence must concern a generic activity, state etc., rather than an individual, or specific one. For instance, since a person dies only once, the death of a specific individual is a unique, non-repeatable event, which would explain the unacceptability of (5.15):

(5.15) *Dan wa, sensoo de shinda koto ga aru 'Dan died in the war'

Presumably, however, the mythical bird Phoenix should be able to say:

(5.16) *Watashi wa, shinda koto ga aru* 'I have died (i.e. experienced dying)'

The use of EXPER seems to be disfavoured even by the mere presence of a definite temporal frame – although it does not necessarily make the event unique. Inoue gives examples with time adverbials denoting periods of time and says that as 'the time span becomes shorter ... the sentence becomes increasingly unacceptable'. We can also see that in the sentences in the questionnaire that contain definite time adverbials, e.g. (Q.141–4), EXPER is normally not used.

On the other hand, it appears that the occurrence of EXPER is favoured by non-affirmative contexts, i.e. questions and negated sentences. Of the top ten examples in table 5.5, five are of this kind. This is natural, since this is where we find non-specific reference to generic events.

The relation between EXPER and PFCT

The Experiential clearly overlaps in its use with the Perfect. As noted above, the Perfect is often said to have several different uses, one of which is referred to as the 'existential reading of the perfect' (McCawley) or the 'experiential perfect' (Comrie). It is pointed out in these treatments that some languages make a formal distinction between this use and other uses of the Perfect. Such formulations suggest that we are in such cases dealing with an opposition between two categories, which between them divide up the domain of the English Perfect. As can be seen from our tables, however, it is not always possible to find a 'non-experiential perfect' in the languages where EXPER occurs.

Among the West Indonesian languages, there are examples of two possible systems. In Sundanese and Indonesian, the non-experiential PFCT and the EXPER markers are mutually exclusive. In Javanese, on the other hand, the EXPER marker *tahu* normally co-occurs with the PFCT marker *uwis*, as in the following sentence (Q.43):

(5.17) *Aku uwis tahu ketemu bapakmu* 'I met your father at least once' (=43)

Furthermore, it is not self-evident that the distribution of EXPER is really included in the distribution of PFCT. We may note that there seems to be a difference between EXPER – at least in its Japanese version – and the English Perfect in that EXPER can sometimes be used when the English Perfect cannot. Inoue compares the unaccep-

tability of (5.18) (given the fact that Faulkner is dead) with the acceptability of the corresponding Japanese sentence (5.19):

(5.18) *Faulkner has written a detective story
(5.19) *Faulkner wa, tantei-shoosetsu o kaita koto ga aru*

The 'repeatability' condition on the Japanese construction does not entail 'repeatability at the moment of speech', as seems to be the case in the English (5.18). It is unclear, however, if this difference depends on idiosyncrasies of the languages involved or upon a real difference between EXPER and PFCT. Examples like (5.18) have been extensively discussed in the literature, and judgements seem to vary (cf. Comrie 1976, 60), which suggests that we are here in a vagueness zone. In the questionnaire, relevant examples would be (Q.43–4) and (Q.50): if we compare their placement on the EXPER and PFCT lists, we see that although there is a clear difference, there is still a considerable number of languages that use a PFCT here. (The examples might have been more decisive if *my father* had been in an unequivocal topic position.)

At one point, there is a very clear and telling distributional difference in our material. It concerns the top sentence on the PFCT list, (Q.53):

(A: I want to give your brother a book to read, but I don't know which. Is there any of these books that he READ already?)
B: (Yes,) he READ this book

One might think that this would be a clear candidate for EXPER: 'He has read the book at some indefinite point in time.' However, we find that only one language uses EXPER here – not unexpectedly, Chinese, which as we have already seen, has a rather deviant EXPER category. The most plausible explanation for this rather striking fact is the tendency we noted above for EXPER to be used primarily in non-assertive contexts. There is a rather similar example, viz. (Q. 37), where we might expect to find EXPER but where again only Chinese has it.

The contexts where EXPER is used seem to coincide at least partly with the ones where the Imperfective of some Slavic languages is used in what is sometimes called the 'general factual' or 'simple denotative' sense (see p. 75), and with a similar use of the Partitive for direct object marking in Estonian and Finnish. Interestingly, however, the use of the latter constructions coincides with a 'Chinese' conception of experientiality rather than the one represented by the other languages where EXPER is found: cases like (Q.53), which we discussed above, are often used as paradigm examples of the 'general factual' use of the Slavic Imperfective. It is possible that both Chinese and Slavic are examples of a category 'non-resultative past' whose distribution might

overlap both with EXPER and the tentative category called FRAMEPAST discussed below.

The Pluperfect (PLPFCT)

On p. 67, we discussed the problems of complex TMA categories. The PLUPERFECT is a paradigm example of these problems. In a language like English, the Pluperfect (as in *John had run*) would appear to be nothing but a combination of the two categories Past and Perfect, as also the term 'Past Perfect' suggests. There are, however, a number of reasons why one would like, in an investigation like this, to treat the PLUPERFECT (PLPFCT) as a separate although not necessarily an independent category. First, there are languages, such as modern spoken French or some dialects of German, which do not have a separate PFCT category but still have a PLUPERFECT. Second, PLUPERFECT may develop secondary or extended uses which are not characteristic of PFCT by itself. Thus, as we shall see below, PLUPERFECT tends to be used for referring to the remote past. Thirdly, it has been noted (as will also be discussed below) that in e.g. English, there are less strict constraints as concerns use with definite time adverbs in the Pluperfect than in the Present Perfect.

In view of these facts, I have chosen a somewhat schizophrenic treatment of PLPFCT. In obtaining the PFCT ranking list, I have normally included both Present Perfects and Pluperfects in the languages where these categories exist as different tenses of the same constructions: this has the disadvantage that Pluperfects are counted twice – both as special cases of PERFECTS and as an autonomous category – but it seemed to be the simplest solution in the complex situation where both tense-marked and non-tense-marked PFCT categories are to be compared.

All instances of PLPFCT, whether definable as past tenses of PFCT or not, are listed in table 5.7.

Table 5.7 PLPFCT categories

Language	Code	Description	Frequency	C	HR
PLPFCT					
Romanian	23	Pluperfect		8	.66 .88
English	H20	Past of <i>have</i> + PtP		11	.67 .91
Italian	A22=	Piuperfetto		10	.62 .80
Kurdish	24	Pluperfect		9	.61 .78
Azerbaijani	22P	<i>-miş</i> + <i>-dir</i>		13	.60 .77
Latin	230	Plusquamperfectum		18	.57 .72
Portuguese	T22	Imperfect of <i>ter</i> + PtP		8	.52 .63
Modern Greek	E20	Past of Aux. <i>écho</i> + Supine		14	.53 .64

Table 5.7 continued

Catalan	H21	Impfct of <i>haver/esser</i> + Pt	15	.50	.60
Limouzi	H22	Imperfect of <i>o (eitre)</i> + PtP	15	.49	.58
Spanish	H21	Imperfect of <i>haber/estar</i> + Ptp	10	.53	.64
Estonian	O2	Past of Aux. <i>olá</i> + PtP	14	.56	.69 <i>olla</i>
Swedish	H2=	Past of Aux. <i>ha</i> + Supine	20	.48	.58
Hindi	A2=	PtP + Past of Aux. <i>hona</i>	40	.48	.63
Persian	Bu2	PtP + Past of Aux. <i>budan</i>	14	.48	.57
French	23	Imparfait of <i>avoir (être)</i> + V	9	.47	.56
Punjabi	A23	93 + 20 of <i>hona</i>	22	.45	.55
Tamil	I02	PtP + Past of Aux. <i>irykka</i>	8	.44	.50
<PLPFCT					
German	H20	Past of <i>haben/sein</i> + PtP	7	.38	.43
Finnish	O2	Past of Aux. <i>olla</i> + Supine	18	.37	.44
Amharic	H2	Plupfct	20	.37	.45
Bengali	A20	PtP + Past (contracted) copula	42	.27	.40
Oromo (Galla)	eT	e= + Aux. <i>ture</i>	11	.05	.00

Table 5.8 Distribution of PLPFCT

Rank no.	No. of categories	Examples
1	20	901 1382
3	19	1392
4	18	891
5	17	481
6	14	491 1291
8	13	521
9	10	511 1061
11	7	671 681 1011
14	6	461 611 1441
17	5	431 451 621 631 1132 1431
24	4	391 1012 1062 1771 1781 1791

Table 5.9 Prototypical occurrences of PLPFCT

Sent. 138 Verb 2
 No. of categories: 20
 When I COME home (yesterday), he WRITE two letters (=he finished writing them just before I came)

Table 5.9 continued

Sent. 90 Verb 1

No. of categories: 20

Q: Did you find your brother at home? A:

(No, we did not, we were very unlucky.) He LEAVE (just before we came)

Sent. 139 Verb 2

No. of categories: 19

When I COME home (yesterday), he WRITE two letters (=that is what he accomplished during my absence)

Sent. 89 Verb 1

No. of categories: 18

Q: Did you find your brother at home? A:

(No, we did not.) He LEAVE (before we arrive)

Sent. 48 Verb 1

No. of categories: 17

Q: When you came to this place a year ago, did you know my brother?

(Yes,) I MEET him (at least once before I came here)

Sent. 49 Verb 1

No. of categories: 14

Q: When you came to this place a year ago, did you know my brother?

(Yes,) I MEET him (just before I came here)

Sent. 129 Verb 1

No. of categories: 14

Looking at a picture of a house which has been torn down

Who BUILD this house?

The prototypical cases of PLPFCT (see tables 5.8–9) are no doubt those that coincide with the traditional definition of the meaning of that category – where one is speaking of an event that took place before a definite point in past time (Reichenbach's R). As suggested above, there are also secondary uses, some of which will be discussed in next section. One secondary use of PLPFCT, well known from English, is in counterfactual constructions (Q.106). As can be seen from table 5.8, this is fairly common – there are ten cases of PLPFCT in the *if*-clause and four of PLPFCT in the main clause of (Q.106).

Marking of past temporal frames

In view of the use of the Present Perfect in English to refer to events within a 'current temporal frame' (see p.136) one would expect that Pluperfect – being the Past of the Present Perfect – should be used to

refer to events within a past temporal frame. However, as was pointed out by McCawley 1971, one says (5.21) rather than (5.20):

(5.20) Henry VIII had been married six times

(5.21) Henry VIII was married six times.

McCawley suggests that there is an asymmetry between the two tenses of the Perfect in English. It should be noted, though, that there is a motivation for such an asymmetry: in the case of the Present Perfect, there is no clear distinction between saying that an event took place before an R and within an F, since if the F is still open at S, S provides an automatic endpoint for it. It now turns out that this is a point where languages may differ: quite a few languages allow the use of a PLUPERFECT in cases like (5.20–1). The comparable sentences in the questionnaire include (Q.129), which contrasts with (Q.128) in that in the latter we are talking about something that still exists.

(Q.128)

(Looking at a house:) Who BUILD this house?

(Q.129)

(Looking at a picture of a house which has been torn down:)
Who BUILD this house?

(Q.129) is not very natural with a Pluperfect in English, but takes a PLPFCT (as the only alternative or alongside with other categories) in about three fourths of the languages that have such a category in our material (Italian, Kurdish, Azerbaijani, Catalan, Bulgarian, Modern Greek, Limouzi, Spanish, Estonian, Hindi, Persian, French, Punjabi, Finnish). As an illustration, here are (Q.128–9) in Modern Greek:

(Q.128:GR) pços éspase aftó tospíti

(Q.129:GR) pços íçe xtísi aftó tospíti

It appears that the use of PLPFCT categories to signal past temporal frames is the first step towards a situation where they are used as a general remote past. Such a development appears to have taken place in several Indic languages, e.g. Hindi-Urdu and Bengali, and also in e.g. Amharic. Cf. the following characterization of the use of the Pluperfect in Hindi from Katenina 1960: it expresses 'the completion of an action before a definite moment (or an action) in the past, and also emphasizes the remoteness of the action from the present moment, its belonging to a finished segment of time – last year, yesterday, yesterday morning etc.' Here – although not in the order they have assumedly developed – we see all the three uses we have been discussing in this section.

In addition, there are a couple of languages in our sample which apparently give 'past temporal frames' special treatment:

(i) Akan appears to have two 'pluperfects', one which is used in the

'normal' cases, such as (Q.138:2), and one which is used in cases like (5.21). Akan has a PFCT (e.g. *óawu*, 'he died') and one category which I have interpreted as a PFV (e.g. *ówui*, 'he died'), which, like most PFVs, is restricted to past time reference. There is also a free morpheme *na* which is used as a past time marker. This marker can occur with both the PFCT and the PFV. Although the data are not quite clear, it appears that the function of *na* with the PFV is to mark that the F is in the past. Otherwise, it is very hard to explain why the PFV may occur both with and without this marker.

(ii) Oneida has a separate morphological category, referred to as the 'perfective past', which morphologically looks as if it were a Past of the PFCT/STAT (see above), but which appears to be used only in the 'past temporal frame' contexts. (In prototypical PLPFCT contexts, PFCT/STAT is used.) Thus, (Q.128-9) in Oneida are as follows:

(Q.128:ON) *uhka?yakonuhsísu* 'Who built (PFCT/STAT) this house?'

(Q.129:ON) *uhka?yakonuhsísu?u* 'Who built (FRAMEPAST) this house?'

The full distribution of the Akan and Oneida 'framepasts' is given in table 5.10. As we can see, the case is clearer for Oneida than for Akan.

Table 5.10 Use of 'framepasts'

Akan	Oneida
511	31
521	321
611	391
621	611
651	1112
1271	1122
1281	1132
1291	1291
1301	
1471	

Interestingly, the PFV:IPFV opposition is used in Russian to make a contrast between (Q.128) and (Q.129):

(Q.128:RU) *Kto postroil ètot dom?* 'Who built (PFV) this house?'

(Q.129:RU) *Kto (v svoe vremja) stroil ètot dom?* 'Who built (IPFV) (at that time) this house?'

It is striking that both Akan and Oneida use the 'framepast' in (Q.61), the classical example of a context where Russian uses an imperfective verb to denote what is in the literature called a 'two-way action', that is, an action which led to a result which has later been cancelled:

(Q.61) (It is cold in the room. The window is closed. QUESTION:)

You OPEN the window (and closed it again)?

(Q.61:AK) *Wo na wobiee mpoma mo anaa?*

(Q.61:ON) *Sawisakalatatú.ne ka?*

(Q.61:RU) *Ty otkryval okno?*

The quotative (QUOT)

QUOTATIVE (QUOT) is a category which appears under many names, such as 'quotative', 'imperceptive', 'non-witnessed', 'second-hand information'. After some hesitation I rather arbitrarily chose the label QUOTATIVE.

QUOT appears in seven languages listed in table 5.11. They seem to be fairly well spread over genetic groups. There are no clear tendencies as to marking type in the QUOT categories: both morphological and periphrastic cases are found.

Table 5.11 QUOT categories

Language	Code	Description	Frequency	C	HR
Beja	S	V + <i>y-d-y</i>	7	.58	.71
Quechua	im	Imperceptive Past	18	.57	.72
Kurdish	23	(Pfct) Inferential	14	.53	.64
Bulgarian	im	Imperceptive Mood	17	.53	.65
Japanese	W	<i>soo da</i> construction	10	.50	.60
Greenl. Eskimo	gs	Suffix <i>sima</i>	8	.44	.50
Turkish	R	Reportative Mood	8	.22	.25

In a number of languages, the functions of QUOT are carried by the PFCT category. We refer to this variety of PFCT as PFCT_q. It is discussed in more detail below.

In general, QUOT marking is most common with past time reference. Our material is too small to make any clear statements about what happens when something non-witnessed with present time reference is reported. (Q.7-8) would be cases of this. As can be seen from them, constructing a natural situation where something of that

'normal' cases, such as (Q.138:2), and one which is used in cases like (5.21). Akan has a PFCT (e.g. *óawu*, 'he died') and one category which I have interpreted as a PFV (e.g. *ówui*, 'he died'), which, like most PFVs, is restricted to past time reference. There is also a free morpheme *na* which is used as a past time marker. This marker can occur with both the PFCT and the PFV. Although the data are not quite clear, it appears that the function of *na* with the PFV is to mark that the F is in the past. Otherwise, it is very hard to explain why the PFV may occur both with and without this marker.

(ii) Oneida has a separate morphological category, referred to as the 'perfective past', which morphologically looks as if it were a Past of the PFCT/STAT (see above), but which appears to be used only in the 'past temporal frame' contexts. (In prototypical PLPFCT contexts, PFCT/STAT is used.) Thus, (Q.128-9) in Oneida are as follows:

(Q.128:ON) *uhka?yakonuhsísu* 'Who built (PFCT/STAT) this house?'

(Q.129:ON) *uhka?yakonuhsísu?u* 'Who built (FRAMEPAST) this house?'

The full distribution of the Akan and Oneida 'framepasts' is given in table 5.10. As we can see, the case is clearer for Oneida than for Akan.

Table 5.10 Use of 'framepasts'

Akan	Oneida
511	31
521	321
611	391
621	611
651	1112
1271	1122
1281	1132
1291	1291
1301	
1471	

Interestingly, the PFV:IPFV opposition is used in Russian to make a contrast between (Q.128) and (Q.129):

(Q.128:RU) *Kto postroil ètot dom?* 'Who built (PFV) this house?'

(Q.129:RU) *Kto (v svoe vremja) stroil ètot dom?* 'Who built (IPFV) (at that time) this house?'

It is striking that both Akan and Oneida use the 'framepast' in (Q.61), the classical example of a context where Russian uses an imperfective verb to denote what is in the literature called a 'two-way action', that is, an action which led to a result which has later been cancelled:

(Q.61) (It is cold in the room. The window is closed. QUESTION:)

You OPEN the window (and closed it again)?

(Q.61:AK) *Wo na wobiee mpoma mo anaa?*

(Q.61:ON) *Sawisakalatatú.ne ka?*

(Q.61:RU) *Ty otkryval okno?*

The quotative (QUOT)

QUOTATIVE (QUOT) is a category which appears under many names, such as 'quotative', 'imperceptive', 'non-witnessed', 'second-hand information'. After some hesitation I rather arbitrarily chose the label QUOTATIVE.

QUOT appears in seven languages listed in table 5.11. They seem to be fairly well spread over genetic groups. There are no clear tendencies as to marking type in the QUOT categories: both morphological and periphrastic cases are found.

Table 5.11 QUOT categories

Language	Code	Description	Frequency	C	HR
Beja	S	V + <i>y-d-y</i>	7	.58	.71
Quechua	im	Imperceptive Past	18	.57	.72
Kurdish	23	(Pfct) Inferential	14	.53	.64
Bulgarian	im	Imperceptive Mood	17	.53	.65
Japanese	W	<i>soo da</i> construction	10	.50	.60
Greenl. Eskimo	gs	Suffix <i>sima</i>	8	.44	.50
Turkish	R	Reportative Mood	8	.22	.25

In a number of languages, the functions of QUOT are carried by the PFCT category. We refer to this variety of PFCT as PFCTq. It is discussed in more detail below.

In general, QUOT marking is most common with past time reference. Our material is too small to make any clear statements about what happens when something non-witnessed with present time reference is reported. (Q.7-8) would be cases of this. As can be seen from them, constructing a natural situation where something of that

kind would be elicited is not quite easy – (Q.7–8) are rather glaring examples of cultural bias.

(Q.7)

Context: A: I just talked to my brother on the phone. B: What he DO right now? A:

Sentence: He WRITE letters

(Q.8)

Context: A: I just talked to my brother on the phone. B: What he DO right now? A:

Sentence: He WRITE a letter

Semantics of QUOT

The typical distribution of QUOT can be seen from tables 5.12–13. The basic function of QUOT is to indicate that the speaker has not experienced himself what he is reporting in the sentence but rather has it from a secondary source. Normally this secondary source is the verbal report of some other person. Probably most languages have some way of indicating that something has been learnt by hearsay, e.g. by constructions such as *be said to* in English, *soll* in German or *lär* in Swedish. These differ from QUOT in that they are normally used only when the speaker wants to emphasize that he does not want to commit himself completely to the truth of the proposition – in other words, when the source of information is relevant to the intended message. QUOT, on the other hand, tends to be systematically used (i.e., exhibits the typical behaviour of a true accident category) as soon as something is told that the speaker has not seen himself. Another difference between the expressions we mentioned and QUOT (most probably also explainable in Gricean terms) is that the former cannot normally be used if the source of the information is indicated in the discourse. Thus, in Text B4, from which most of the examples in table 5.13 are taken, one could not use *be said to* in English, *soll* in German or *lär* in Swedish – if one did, it would imply that the source would not be my brother. The same holds true of the so-called ‘modus obliquus’

Table 5.12 Distribution of QUOT

Rank no.	No. of categories	Examples
1	5	1771 1781 1792 1801
5	4	1341 1761
7	3	1361 1791 1811 1851

Table 5.13 Prototypical occurrences of QUOT

Sent. 177 Verb 1 No. of categories: 5

Do you know what happened to my brother yesterday? He told it ... (narr). ...
Suddenly he STEP on a snake

Sent. 178 Verb 1 No. of categories: 5

(Context: see above)
It BITE him in the leg

Sent. 179 Verb 2 No. of categories: 5

(Context: see above)
He TAKE a stone and THROW at the snake

Sent. 180 Verb 1 No. of categories: 5

(Context: see above)
It DIE

Sent. 134 Verb 1 No. of categories: 4

A person who has heard (133) but not seen the event says:
The king ARRIVE

Sent. 176 Verb 1 No. of categories: 4

Do you know what happened to my brother yesterday? He told it himself.
He WALK in the forest

in Estonian (marked by the suffix *-vat*), which otherwise seems to be a rather systematic marker of second-hand information, cf.:

(Q.133:ES) Kuningas saabus ‘The king has arrived (Past indicative)’

(Q.134:ES) Kuningas olevat saabunud ‘The king has (so they say) arrived (Modus obliquus)’

There are some important differences between the languages in table 5.11 as to use of QUOT.

In some languages – Bulgarian, Turkish and Quechua – our informants use QUOT in text B5, which is intended to suggest a fictional context. In Bulgarian, QUOT is often used in story-telling, e.g. in folk-tales. (In literary fiction, QUOT apparently tends not to be used – see Stojanov 1964, 389.) A similar situation seems to hold in Turkish (Slobin & Aksu 1982).

In Bulgarian, QUOT is also used to mark indirect speech in subordinate clauses. This is not the case in the other QUOT languages.

A speaker may make a statement about something that he has not witnessed about which he makes an inference on the basis of things he knows or can perceive. (Q.60) is an illustration of this:

(Q.60)

Context: Investigating a burglary, seeing footprints beneath a window:

Sentence: The thief ENTER the house by this window

It should be noted that this is different from the typical use of QUOT discussed above, where the source of information is another person. In some languages with QUOT, but not all, QUOT is used in contexts like (Q.60), viz. Quechua and Turkish. In Bulgarian, on the other hand, PFCT is used here. This shows that the label 'inferential' used e.g. in Comrie 1976 is not suitable for QUOT, since it implies that there is no distinction between cases like (Q.60) and hearsay cases such as (Q.134).

The relation between PFCT and QUOT

As was mentioned above, PFCT is in some languages used in the contexts where QUOT appears. The categories in question (referred to as PFCTq) are listed in table 5.1. In our material, this looks like an areal phenomenon: its boundaries coincide fairly well with the area of the Ottoman empire. It seems to be documented also in a number of Uralic languages in the Soviet Union, however (Comrie 1976, 110).²

Within the 'Ottoman' area, there appears to have been a historical development, which has passed through several stages although it has not gone equally far in all languages. The first step seems to have been an extension of an old PFCT to quotative use: the stage which this resulted in is still represented by the PFCTq languages. From here, at least two different developments are possible: either the primary uses of the old PFCT category are taken over by some other category – this has taken place in Turkish and Kurdish – or the PFCTq category splits up into two: this is apparently what has happened in Bulgarian, where PFCT and QUOT are now distinguished at least in the third person by the absence of the copula in the QUOT form. In Turkish and Bulgarian, this development has been accompanied by the development of QUOT categories also for non-PFV forms; in Kurdish, the only such category is the Quotative Imperfect.

It seems not uncommon for PFCT to have secondary 'inferential' or 'quotative' uses. In English, this seems particularly typical of the Perfect Progressive, as in the English version of (Q.59):

(Q.59:EN) It has been raining

In written Swedish, particularly in more formal styles, a quotative use of the Perfect can be found:

(5.22) Vittnet har lämnat lokalen klockan två '(lit.) The witness (says that he) has left the premises at two o'clock'

Interestingly, there seems to be no reluctance against the use of definite time adverbials here (in contradistinction to 'normal' uses of the Perfect in Swedish).

'Inferential' uses can be regarded as a rather natural extension of the primary uses of PFCT: the inferential involves making a conclusion about a past event on the basis of its remaining effects, something which will also quite often be true of 'the perfect of result' (e.g. when one says *He has left* looking into someone's empty office). It is somewhat harder to see a direct link between the quotative use and the prototypical uses of PFCT. My suggestion is that the quotative use is tertiary in the sense that it has developed as an extension of the secondary inferential use.

6

An overview of the TMA systems of the languages in the sample

In this chapter, we shall survey the TMA systems of the languages in the sample, sorted by genetic groups. For each language, a list of its 'major TMA categories' is given, with indications of their codes in the data base, their traditional names (for morphological categories) or their constituents (for periphrastic constructions or morphological categories which lack a traditional name), their frequencies in the questionnaire, their analyses in terms of cross-linguistic categories, and their marking-types. In addition, sample paradigms are provided for a number of languages.

For each language group, we list the languages for which sketches can be found in Dahl & Kós-Dienes 1984.

(AA) Afro-Asiatic

(AA1) Cushitic

Table 6.1 Major TMA categories: Cushitic

Code	Description	Analysis	Frequency	Marking type
Beja				
10	Present	DEFAULT	40	M
21	Preterite	PFV	94	M
22	Past	PASTid	10	M
C	PtP + enclitic copula	<PFCT	12	P
S	V + <i>y-d-y</i>	QUOT	7	P
Y	Imperative + <i>y-d-y</i>	FUT	6	P
Oromo (Galla)				
a=	Imperfetto	DEFAULT	40	M
aG	Imperfetto + Aux. <i>gira</i>	PRES-PROG	11	P
T	Aux. <i>ture</i>	PASTi	45	P

Table 6.1 continued

e=	Perfetto	<PFV	85	M
eT	Perfetto + Aux. <i>ture</i>	PLPFCT	11	P
er	Passato prossimo	<PFCT	8	M
<i>Beja paradigms: tam- 'eat'; d-b-l- 'collect'</i>				
10	tam-fi-ni	danbíil		
21	tam-y-à	i-dbíil		
22	tam-i	iidbíil		
C	tam-aa-'b-u			

This group is represented by two languages, Beja and Oromo (Galla). Beja has a system of verb morphology reminiscent of the Semitic languages with complicated patterns of vowel alternations. Oromo morphology is much simpler, relying mostly on suffixes. Beja is notable for having PAST marking only in the dynamic subsystem.

Sketch: Beja.

(AA2) Semitic

Table 6.2 Major TMA categories: Semitic

Code	Description	Analysis	Frequency	Marking type
Modern Standard Arabic				
1	Imperfective	DEFAULT	84	M
2	Perfective	PFV	124	M
3	Future	FUT	13	M
B	<i>bişasadi</i> + Verbal Noun	PROG?	6	P
K1	Copula + Imperfective	PASTi	21	P
Arabic (Tunisian)				
K	<i>ka:n</i> + Impfct/Adj.	PASTi	17	P
i	Imperfect (Present)	DEFAULT	95	M
p=	Perfective (Past)	PFV	128	M
Hebrew (Modern)				
10	Present	DEFAULT	64	M
20	Past	PAST	141	M
30	Future	FUT	30	M

Table 6.2 continued

Maltese				
K	<i>kien</i> + Imperfective	PASTi	45	P
Q	<i>qieghed</i> + Imperfective	PROG	19	P
S	<i>ser</i> + Imperfective	FUT	6	P
i	Imperfective	DEFAULT	119	M
p=	Perfective	PFV	106	M
Tigrinya (Tigrigna)				
10	Present	DEFAULT	35	M
30	Future	FUT	12	M
92	Simple Gerund	PAST	149	M
92E	92 + Past Copula	<PASTn	9	P
pr	Progressive	PRES-PROG	7	M
Amharic				
1=	Present	DEFAULT	82	M
2=	Simple Past	PAST	92	M
H1	Perfect	PFCT	19	M
H2	Gerund + Aux. <i>nebbet</i>	PLPFCT	20	P
N2	Present + Aux. <i>nebbet</i>	PASTi	10	P

This family is relatively well represented. Three of the six representatives are so closely related, however, that they should perhaps rather be called dialects of one language: there are Modern Standard Arabic (i.e. the variety of Classical Arabic used as an official language in most Arabic countries); Tunisian Colloquial Arabic – one of the many spoken varieties of Arabic; and Maltese, which differs from the last-mentioned group by being also a written language in Malta. There are enough differences in the TMA systems, though, to merit inclusion of all these in the sample.

All the Arabic dialects share the typical Semitic verb system with extensive use of stem alternations. There is a basic distinction between PFV and IPFV verb forms, characterized by the position of the subject agreement marker. There is also a periphrastic PASTi category, which may have become more systematically used in recent dialects than it was originally.

The TMA system of Modern Hebrew resembles the Arabic one as to the morphological means used, but there seems to have been a major shift in the function of the main TMA categories in post-Biblical times, there now being no PFV:IPFV but only PAST.

Also the TMA systems of Amharic and Tigrinya (classified as belonging to the Ethiopic or African Semitic subgroup) seem rather different from the Arabic one.

For a sample paradigm, see p. 83.

Sketches: Tunisian Arabic, Maltese, Tigrinya.

(AC) Altaic

(ACI) Turkic

Table 6.3 Major TMA categories: Turkic

Code	Description	Analysis	Frequency	Marking type
Azerbaijani				
11	Present	IPFV	49	M
21	Categorical Past (-dl)	PFV	46	M
22	Suffix -mlsh	PFCTq	68	M
22P	22 + P	PLPFCT	13	M
31	Categorical Future	FUT	12	M
32=	Non-categorical Fut., Aorist	PRED	8	M
32P	32 + P	HABPASTc	5	M
D	Pres. Predicative Affix (-dlr)	DEFAULTs	15	M
P	Past Pred. Affix (-dl)	PASTi	46	M
Turkish				
110	Suffix -yor	IPFV	48	M
120	Aorist	HABG/PRED	24	M
210	Past Definite	PFV	70	M
30	Future	FUT	13	M
Nr	Past Predicative Suffix	PASTi	21	M
R	Reportative Mood	QUOT	43	M
Azerbaijani paradigm: öl- 'die'				
11	ölür			
21	öldü			
22	ölmüşdür			
22P	ölmüsdü			
31	öləcək			
32=	ölür			
32P	ölürdü			

From the Turkic group we have two languages, Turkish and Azerbaijani. These have rather similar TMA systems, although the old PFCT which is still PFCTq in Azerbaijani has moved to a QUOT in Turkish. There is also a difference between the Aorist in Turkish which we have analysed as HABG/PRED and the historically

corresponding form in Azerbaijani which seems more like a pure PRED. These two situations seem to represent different stages in a development whereby an old PROG (forms with the *-yor* suffix) has developed into a general IPFV or default category. Notice, however, that the combination of PAST and PRED in Azerbaijani yields a form used as a HABPAST.

(AC2) Japanese

Table 6.4 Major TMA categories: Japanese

Code	Description	Analysis	Frequency	Marking type
1=	Present	DEFAULT	37	M
2=	Past (form)	<PFV	41	M
D	<i>daroo</i>	PRED	13	P
I	<i>-te iru</i> construction	IPFV/RESUL	80	P
Kg	<i>koto ga aru</i> construction	EXPER	8	P
Sh	<i>-te shimau</i> construction	CONCL	18	P
W	<i>soo da</i> construction	QUOT	10	P

Japanese has a fairly complex TMA system. The main aspectual distinction is between the forms that involve the *-te iru* construction and the simple verb forms: the *-te iru* forms have a complex semantics and apparently correspond to two different categories in older stages of the language and some non-standard dialects (Martin 1975). The analysis we give – IPFV/RESUL – must be regarded as tentative.

(AD) Andean-Equatorial

Table 6.5 Major TMA categories: Andean-Quechumaran

Code	Description	Analysis	Frequency	Marking type
Quechua				
20	Past (suffix <i>-rqa-</i>)	PASTn	75	M
3=	Future	FUT	12	M
U	Unmarked	DEFAULT	98	M
im	Imperceptive Past (suffix <i>-sqa-</i>)	QUOT	18	M
pr	Progressive (suffix <i>-sa-</i>)	PROG/STAT	51	M

Table 6.6 Major TMA categories: Tupi

Code	Description	Analysis	Frequency	Marking type
Guarani				
1=	Unmarked	DEFAULT	108	U
20'	Past (suffix <i>-kuri</i>)	PASTn	117	M
30	Future (suffix <i>-ta</i>)	FUT	12	M
e	Emphatic (suffix <i>-hina</i>)	<PROG	10	M

The two languages from this group, Quechua and Guarani, are rather distantly related but have in common the feature of not marking PAST in narrative contexts. (For a more extensive discussion, see chapter 4.) Both languages, in particular Quechua, have rather complex systems of verbal morphology.

Influence from Spanish is likely in both these languages, which makes it more difficult to evaluate their apparent peculiarities.

Sketches: Guarani, Quechua.

(AN) Australian

Table 6.7 Major TMA categories: Australian

Code	Description	Analysis	Frequency	Marking type
Alawa				
10	Present Continuous	DEFAULT	68	M
21	Past Continuous	PASTi	55	M
22	Past Punctiliar	PFV	97	M
30	Future	FUT	34	M
U	Unmarked	DEFAULTs	16	U
g	Suffix <i>-gay</i>	HABPAST	8	M
Bandjalang				
1	Present	DEFAULT	133	M
21	Past Definite	PFV	84	M
22	Past Indefinite	PASTi+EXPER	78	M
3	Future	FUT	34	M
a	Cont./Antipassive	PROG	50	M

Table 6.7 continued

Alawa paradigm: jangarl-winga 'die, be sick'

10	jangarl-narla 'he is dying/is sick'
21	jangarl-nawujala 'he was sick'
22	jangarl-nayiman 'he died'
30	jangarl-nawinja 'he will die'
21g	jangarl-nawujalagay 'he used to be sick'

From the Australian continent there are two languages in the sample, Alawa, which is spoken in Arnhem Land and classified as a Maran language, and Bandjalang, spoken in New South Wales and belonging to the Pama-Nyungan group. (Both Maran and Pama-Nyungan are subgroupings within the Australian phylum.) Their TMA systems seem rather similar to each other, although there are some unclear points in the analysis. Thus, the categories 'Definite Past' and 'Indefinite Past' respectively in Bandjalang have here been analysed as PFV:IPFV, although the correlation figures are not too impressive: there seems to be quite considerable free variation between the two categories. In both languages, marking of TMA categories is exclusively suffixal.

(AU) Austronesian

(AUI) West Indonesian

Table 6.8 Major TMA categories: West Indonesian

Code	Description	Analysis	Frequency	Marking type
Cebuano				
bs	Base	DEFAULT	9	U
rdu	Prefix <i>ga-</i>	IPFV	48	M
rpo	Prefix <i>naka-</i>	PFV N-VOL	25	M
rv	Prefix <i>mi-</i>	PFV VOL	18	M
u=du	Prefix <i>mag-</i>	<PRED	9	M
u=po	Prefix <i>maka-</i>	<EXPER	11	M
u=v	Prefix <i>mag-</i>	<HABG	20	M
Indonesian				
A	<i>akan</i> + V	FUT	13	P
P	<i>pernah</i> + V	EXPER	12	P
S=	<i>sudah</i> + V	PFCT	21	P
SE	<i>sedang</i> + V	PROG	17	P
U	Unmarked	DEFAULT	181	U

Table 6.8 continued

Javanese

A	<i>arep</i> + V	FUT	9	P
L	<i>lagi</i> + V	PROG	22	P
T	<i>tahu</i> + V	EXPER	8	P
U	Unmarked	DEFAULT	167	U
W	(<i>u</i>) <i>wis</i> + V	PFCT	24	P

Sundanese

B=	<i>bade</i> + V	PROSP	6	P
K=	<i>kantos</i> + V	EXPER	10	P
N=	<i>nuju</i> + V	PROG	17	P
Pa	<i>parantos</i> + V	PFCT	23	P
T	<i>tadi</i> + V	<PAST	6	P
U	Unmarked	DEFAULT	176	U

Of the four languages in this group, Indonesian, Javanese and Sundanese are closely related and have relatively similar TMA systems, which are characterized by the complete absence of morphologically expressed categories. In particular, all these languages lack both PFV:IPFV and PAST. On the other hand, they all have PFCT, EXPER and PROG. It should be noted, however, that the languages of this group have well developed (morphologically expressed) voice systems, which apparently may express aspectual notions (see e.g. Hopper 1982a and Rafferty 1982 for discussion), although this is only weakly reflected in our material, something which is probably at least partly due to the fact that we have not tried to elicit non-active constructions systematically.

The fourth language, Cebuano, is probably the most difficult language in the sample to describe on the basis of the questionnaire. The TMA system is linked up in the morphology with a well-developed voice system and an opposition between 'volitional' and 'non-volitional' ('potential' in the terminology of Wolff 1966) action, which complicates the matter considerably. The verb morphology is mainly based on a set of prefixes, which are 'portmanteau', i.e. express at once several morphological categories. For this reason, it is difficult to judge how many dimensions should be postulated in the morphological system.

In general, we have arrived at rather low correlations with the assumed 'ideal' distributions of categories. The identifications we have made below are therefore rather disputable.

It seems that PFV:IPFV is realized as an opposition between the imperfective 'Durative' (or probably rather the 'Real Durative' with the prefix *ga-*) and the other (perfective) 'Real' forms, i.e. the 'Real Potential' (prefix *naka-*), the 'Real Volitional' (prefix *mi-*), and the

different passives. The 'Unreal Potential' (prefix *maka-*) has a .39 correlation with EXPER. The other 'Unreal' forms usually have either future or habitual-generic (non-past) uses. The 'Unreal Durative' (prefix *ga-*) appears to be mainly used for predictions, and the 'Unreal Volitional' (prefix *mag-*) for habitual-generic cases.

The TMA systems of this group of languages certainly deserve further study.

Sketches: Indonesian, Javanese, Sundanese.

(AU2) *South Celebes*

Table 6.9 Major TMA categories: South Celebes

Code	Description	Analysis	Frequency	Marking type
Bugis Makassar				
A	Adjective	DEFAULTs	28	U
E	<i>ero</i> + V	FUT	11	P
L	<i>leba</i> + V	PFCT	28	P
U	Unmarked	DEFAULT	173	U

The only example from this group, Bugis Makassar, may be used as an illustration of a very simple TMA system, where the majority of the verbs in the questionnaire – more exactly about 200 – receive no marking at all. Thus, (6.1) is used as the translation of (Q.5, 7, 9, 11, 13, 18, 20, 24, 25, 26):

(6.1) Annulisi ki sura'-sura' 'He writes/is writing etc. letters'
write he letters

There are, however, a few particles that may be used to mark TMA categories. *Leba* is used as a PFCT, e.g.:

(Q.37:BG) *Leba ka sibuntulu ki* 'I have met him'
PFCTI meet him

Ero is used as a FUT marker:

(Q.15:BG) *Ero ki annulisi sere' sura'* 'He will write a letter'
FUThe write one letter

(There is also another particle, *jadi*, which also seems to be used for future time reference.)

Sketch: Bugis Makassar.

(AU3) *Polynesian*

Table 6.10 Major TMA categories: Polynesian

Code	Description	Analysis	Frequency	Marking type
Hawaiian				
U	Unmarked	DEFAULT	122	U
i=	<i>e</i> + V + <i>ana</i>	IPFV?	29	P
p=	<i>ua</i> + V	PFV	132	P
pr	<i>ke</i> + V + <i>nei</i>	PRES-PROG	25	P
Maori				
2	<i>i</i> + V	PASTn	51	P
4	<i>e</i> + V	DEFAULT?	20	P
6	<i>ka</i> + V	DEFAULT?	68	P
91	Unmarked stem	?	9	U
a	<i>e</i> + V + <i>ana</i>	PROG	35	P
c	<i>te</i> + V	PROG	19	P
h	V + <i>ai</i>	HABG	8	P
pf	<i>kua</i> + V	PFCT	27	P
r	Reduplicated Verb	IPFV?	51	M

There are two languages from this group in the sample: Hawaiian and Maori. They both lack morphological TMA categories, i.e. the TMA systems in these languages are expressed by periphrastic categories only.

In Hawaiian, the categories interact in sometimes unexpected ways. For instance, there is a seemingly rather mysterious category *e* + V + *ana* which appears to be ambiguous between being 'past progressive' and 'future'. The explanation is probably that the category in question basically represents IPFV, excluding the contexts in which the PRES-PROG *ke* + V + *nei* is found and the generic and habitual contexts where the bare verb is used.

The Maori questionnaire is based on several dialects. Therefore it has happened quite often that several alternative translations are given for the same sentence, and it is not clear if all the categories attested in the questionnaire can be used in the same idiolect, i.e. if they are really coexistent in the same system.

Sketches: Hawaiian, Maori, Tahitian.

(CN) Caucasian

Table 6.11 Major TMA categories: Caucasian/Kartvelian

Code	Description	Analysis	Frequency	Marking type
Georgian				
10	Present	DEFAULT	107	M
21	Imperfect	PASTi	38	M
22	Aorist	PFV	96	M
30	Future	FUT	41	M
X	Particle <i>xolme</i>	HAB	12	P
p	Perfect	PFCTq	45	M

Georgian paradigm: kudoma 'die'

10	ḱudeba 'he is dying'
21	ḱudeboda 'he was dying/died'
22	moḱuda 'he died'
30	moḱudeba 'he will die'
10p	momḱudara 'he has died/died (reportedly)'

There is just one representative of the Caucasian languages: Georgian. The most noteworthy feature of the TMA system of Georgian is that it involves what looks like a 'Slavic-like' aspectual category, the status of which is discussed in chapter 3.

(DR) Dravidian

Table 6.12 Major TMA categories: Dravidian

Code	Description	Analysis	Frequency	Marking type
Tamil				
1=	Present	DEFAULT	32	M
2=	Past	PAST	84	M
3=	Future	FUT/HAB?	28	M
I0	PtP + Aux. <i>irykka</i>	PFCT	24	P
V	PtP + Aux. <i>vida</i>	CONCL	30	P
Z	PrP + Aux. <i>irykka</i>	PROG	29	P
I02	PtP + Aux. <i>irykka</i>	PLPFCT	8	P

Dravidian, the dominating language-family in the southern part of the Indian subcontinent, is represented by Tamil. Tamil has a fairly large number of periphrastic TMA constructions in addition to the morphologically marked ones. A rather problematic category is the one with the auxiliary *vida*, which we have here analysed as CONCL (see chapter 3).

Sketch: Tamil.

(EA) Eskimo-Aleut

Table 6.13 Major TMA categories: Eskimo-Aleut

Code	Description	Analysis	Frequency	Marking type
Greenlandic Eskimo				
2	Suffix <i>-sima-</i>	<PFCT	27	M
30	Suffix <i>-saa-</i>	FUT	15	M
U	Indicative	DEFAULT	85	U
gs	Suffix <i>-sima-</i>	QUOT	8	M
ri	Suffix <i>-riataar-</i>	CONCL?	8	M
t=	Suffix <i>-tar-</i>	HABG	28	M

Greenlandic Eskimo paradigm: tikit- 'come'

U	tikippoq 'he comes/came/is coming'
2	tikissamavoq 'he came/has come'
30	tikissaaq 'he will come'
gs	tikissimavoq 'he comes/came (reportedly)'
ri	tikeriataarpoq 'he suddenly came'
t=	tikittarpoq 'he comes (usually)'

This family is represented by West Greenlandic Eskimo, which has a rather complicated verbal morphology, but which nonetheless does not mark the core categories PFV:IPFV or PAST in any systematic way.

Sketch: West Greenlandic Eskimo.

(IE) Indo-European
(IE1) Germanic

Table 6.14 Major TMA categories: Germanic

Code	Description	Analysis	Frequency	Marking type
Afrikaans				
1=	Unmarked	DEFAULT	127	U
2=	Past copula <i>was</i>	PASTs	16	M
H1	Aux. <i>het</i> + PtP	PASTd	145	P
S1	Aux. <i>sal</i> + Inf.	FUT	23	P
English				
10	Present	DEFAULT	107	M
20	Past	PAST	102	M
BG	<i>be</i> + Gerund	PROG	39	P
H	<i>have</i> + PtP	PFCT	37	P
W	<i>will</i> + Inf.	FUT	20	P
H20	Past of <i>have</i> + PtP	PLPFCT	11	P
Fitzroy Crossing Kriol				
10	Unmarked	DEFAULT	94	U
B	Aux. <i>bin</i> + V	PAST	141	M
G	Aux. <i>garra</i> + V	FUT	12	M
d	Suffix <i>-bat</i>	PROG	41	P
pr	Suffix <i>-in(g)</i>	?	14	P
German				
10	Present	DEFAULT	133	M
20	Past	PAST	95	M
G	Particle <i>gerade</i>	PROG	7	P
H	Aux. <i>haben/sein</i> + PtP	PFCT	64	P
W=	Aux. <i>werden</i> + Inf.	FUT	17	P
H20	20 of <i>haben/sein</i> + PtP	<PLPFCT	7	P
Swedish				
10	Pres	DEFAULT	95	M
2=	Past	PAST	134	M
H	Aux. <i>ha</i> + Supine	PFCT	61	P
K	<i>kommer</i> + att + Inf.	PRED	9	P
S	Aux. <i>skola</i> + Inf.	FUT	12	P
X	<i>hålla på</i> att + Inf.	PROG	16	P
H2=	2= of Aux. <i>ha</i> + Supine	PLPFCT	20	P

The sample includes one Scandinavian language, Swedish, and four West Germanic ones – the familiar ones, English and German (a North German variety), but also two languages representing developments in non-European settings, viz. Afrikaans, one of the languages of the white minority of

South Africa, and Fitzroy Crossing Kriol, an English-based Creole spoken by a group belonging to the aboriginal population of Australia.

The Germanic languages are in general characterized by the lack of a PFV:IPFV distinction. English and Swedish have manifestations of PFCT: Afrikaans represents a development also exemplified by South German dialects and Yiddish, where the old PFCT has been generalized to a PAST, the old PAST being kept only for the copula. The Northern German dialects linger behind, having a PFCT that may be better characterized as a non-narrative PAST.

Being a Creole language, Fitzroy Crossing Kriol is of special interest, since the TMA systems of Creole languages have been claimed to have rather peculiar properties. It seems that Fitzroy Crossing Kriol does not fit e.g. Bickerton's (1981) views of what a Creole TMA system should look like very well: we do not have data enough about the language to judge about its exact status, however, and it may be rash to use it as counterevidence to his claims. It may be noted, however, that Fitzroy Crossing Kriol has rather radically restructured the English TMA system, abolishing the old morphological markings and introducing new, mainly periphrastic constructions.

Sketches: Afrikaans, Fitzroy Crossing Kriol.

(IE2) Indo-Iranian

Table 6.15 Major TMA categories: Indo-Iranian

Code	Description	Analysis	Frequency	Marking type
Bengali				
10	Present	DEFAULT	85	M
20	Past	PAST	81	M
25	Past Frequentative	HABPASTc	13	M
30	Future	FUT	29	M
A	PtP + (contracted) cop.	<PFCT	80	P
T	PrP + (contracted) cop.	PROG	38	P
A20	PtP + Past (contracted) cop.	<PLPFCT	42	P

Table 6.15 continued

Hindi/Urdu				
1=	Present	DEFAULTs	81	M
21	PtP (pred.)	PFV	56	M
2=	Past	PASTs	47	M
3=	Future	FUT	30	M
A	PtP + Aux. <i>hona</i>	<PFCT	63	P
R	Inf. + PtP of <i>rahna</i> + Aux. <i>hona</i>	PROG	31	P
T	PrP + Aux. <i>hona</i>	DEFAULTd	35	P
A2=	PtP + Past of Aux. <i>hona</i>	<PLPFCT	40	P
Kurdish				
10	Pres	DEFAULT	52	M
21	Simple Past	PFV	88	M
22	Past Continuous	PASTi	16	M
23	(Perfect) Inferential	QUOT	14	P
24	Pluperfect	PLPFCT	9	P
30	Future	FUT	27	M
Punjabi				
10	Present	DEFAULTs	85	M
20	Imperfect	PASTs	32	M
21	PtP (predicative)	PFV	74	M
30	Future	FUT	30	M
A	PtP + Aux. <i>hona</i>	PFCT	51	P
A23	PtP + 20 of <i>hona</i>	PLPFCT	22	P
R	V + A of <i>rahna</i>	<PROG	21	P
T10	PrP + Pres. of <i>hona</i>	DEFAULTd	27	P
T20	Inf. + Imperfect of <i>hona</i>	PASTi	11	P
Persian				
10	Present	DEFAULT	100	M
2=	Past	PFV	79	M
Bu	PtP + Aux. <i>budan</i>	PFCTq	48	P
D	Aux. <i>dastan</i> + Verb	PROG	26	P
M2	Imperfect	PASTi	26	M
X	Aux. <i>xastan</i> + Inf.	FUT	16	P
Bu2	PtP + Past of Aux. <i>budan</i>	PLPFCT	14	P

There are two Iranian languages, Kurdish and Modern Persian, and three Indic languages, Bengali, Hindi/Urdu and Punjabi, in the sample.

The two Iranian languages have quite similar systems, although they differ in details. From the morphological point of view, they are noteworthy for having a peculiar realization of the common tripartite PFV:PASTi:DEFAULT system, where the IPFV forms differ from the PFV by having a prefix and the Present (DEFAULT) differs from the two others by having a different stem and/or endings. In these languages, we also find different varieties of the development PFCT → QUOT discussed in chapter 5.

Of the Indic languages, Hindi/Urdu and Punjabi are quite closely related. (We have obtained one questionnaire each from speakers labelling their language 'Hindi' and 'Urdu', respectively: there seemed to be no ground for treating them as two different languages, however.) Their TMA systems are characterized by the extensive use of periphrastic constructions. The system of Bengali is rather different. One peculiarity of this language is the use of the old Pluperfect as a remote past.

Sketch: Kurdish.

(IE3) Greek

Table 6.16 Major TMA categories: Greek

Code	Description	Analysis	Frequency	Marking type
Greek (Modern)				
10	Present	DEFAULT	80	M
20	Past	PAST	156	M
E	Aux. <i>écho</i> + Supine	<PFCT	21	P
T	Particle <i>tha</i>	FUT	23	P
p	Perfective aspect	PFV	104	M
E20	20 of Aux. <i>écho</i> + Supine	PLPFCT	14	P

The TMA system of Modern Greek differs in some respects from that of Classical Greek, which is discussed in chapter 3. Modern Greek is peculiar in having a PFV/IPFV distinction in the future, something which is otherwise found mainly in the 'Slavic' systems.

(IE4) Romance

This group is well represented – maybe a better word would be over-represented – in the sample. In addition, one of the languages, Latin, has the status of a common proto-language (or as close to it as you can get), and although we do not, for natural reasons, have access

Table 6.17 Major TMA categories: Romance

Code	Description	Analysis	Frequency	Marking type
Catalan				
10	Present	DEFAULT	109	M
21	Imperfect	PASTi	50	M
22	Simple Past	<NARR	10	M
30	Future	FUT	17	M
H	Aux. <i>haver/esser</i> + PtP	PFCT	56	P
J	Aux. <i>acabar</i> + <i>de</i> + Inf.	IMMED-PAST	8	P
P	Aux. <i>estar</i> + PrP	<PROG	11	P
V	<i>va</i> + Inf	PFV	56	P
H21	21 of <i>haver/esser</i> +PtP	PLPFCT	15	P
French				
10	Présent	DEFAULT	60	M
21	<i>avoir (être)</i> + V	PFV	98	P
22=	Imparfait	PASTi	38	M
30	Futur Simple	FUT	16	M
T	<i>être en train de</i> + Inf.	PROG	12	P
23	22 of <i>avoir (être)</i> +V	PLPFCT	9	P
Italian				
10	Presente	DEFAULT	174	M
22=	Imperfetto	PASTi	42	M
30	Futuro	FUT	27	M
A	Passato prossimo	PFV	112	P
S	Aux. <i>stare</i> + gerundio	PROG	16	P
A22=	Piuperfetto	PLPFCT	10	P
Latin				
100	Praesens	DEFAULT	52	M
210	Perfectum	PFV	109	M
220	Imperfectum	PASTi	28	M
230	Plusquamperfectum	PLPFCT	18	M
31	Futurum	FUT	21	M
Limouzi				
10	Present	DEFAULT	80	M
21	Past Simple	PFV	85	M
22=	Imperfect	PASTi	38	M
30	Future	FUT	17	M
Ea	<i>eitre a</i> + Inf.	PROG	9	P
H	<i>o (eitre)</i> + PtP	PFCT	54	P
T	<i>eitre en tren de</i> + Inf.	PROG	22	P
H22	22 of <i>o (eitre)</i> +PtP	PLPFCT	15	P

Table 6.17 continued

Romanian (Rumanian)

10	Present	DEFAULT	83	M
21	Simple Past	<PFV	11	M
22	Imperfect	PASTi	31	M
23	Pluperfect	PLPFCT	8	P
A	Aux. <i>avea</i> + PtP	PFV	88	P
O	Aux. <i>o</i> + Subj.	FUT	14	P
V	Aux. <i>voi</i> + Inf.	FUT	14	P

Spanish

10	Present	DEFAULT	114	M
21	Imperfect	PASTi	54	M
22	Simple Past	PFV	65	M
30	Future	FUT	10	M
A	Aux. <i>acabar</i> + <i>de</i> +Inf.	IMMED-PAST	8	P
F	Aux. <i>ir</i> + <i>a</i> + Inf.	<FUT	5	P
H	Aux. <i>haber/estar</i> + PtP	PFCT	55	P
P	Aux. <i>estar</i> + PrP	PROG	37	P
H21	Imperfect of <i>haber/estar</i> + PtP	PLPFCT	14	P

Portuguese

1=	Present	DEFAULT	42	M
21	Simple past	PFV	108	M
22	Imperfect	PASTi	46	M
3=	Future	<FUT	8	M
E	<i>estar</i> + <i>a</i> + Inf.	PROG	24	P
I	<i>ir</i> + <i>a</i> + Inf.	FUT	9	P
T	<i>ter</i> + PtP	<PFCT	12	P
T22	Imperfect of <i>ter</i> + PtP	PLPFCT	8	P

to the intuitions of a native speaker in this case, we are still able to get glimpses of how TMA systems may change in a group of related languages over two millennia. In Latin, the old Indo-European PFCT has taken over the territory of the original PFV (or Aorist): at a later stage – represented in the sample by Spanish, Limouzi, and maybe Portuguese – a new periphrastic construction, the *passé composé* appears, which in some languages (French, Italian) repeats the expansion process, conquering the PFV territory again from the *passé simple*. Romanian seems to be an intermediate case, where the Simple Past is used at least in certain contexts. In Catalan, a rather original periphrastic construction has appeared which competes with the ordinary Simple Past.

(IE5) Slavic

Table 6.18 Major TMA categories: Slavic

Code	Description	Analysis	Frequency	Marking type
Bulgarian				
10	Present	DEFAULT	67	M
20	Aor.=Impfct (for Cop. only)	PASTs	11	M
21	Aorist	PFV	88	M
22	Imperfect	PASTi	22	M
24	Copula + Aorist PtP	PFCT	35	P
30	šte + Present	FUT	20	P
im	Imperceptive Mood	QUOT	17	M
p	Perfective verb	PFVd	106	M
Czech				
10	Present	DEFAULT	126	M
20	Past	PAST	146	P
B	Future (Imperfective)	FUTi	9	P
M	mít + Past Pass. Ptcple	<PFCT	6	P
it	Iterative Verb	HAB	9	M
p	Perfective Verb	PFVd	130	M
Polish				
10	Present	DEFAULT	99	M
2	Past	PAST	153	M
3=	Aux. być + Past/Inf.	FUTi	13	P
p=	Perfective Verb	PFVd	136	M
Russian				
1=	Present (Non-Past)	DEFAULT	79	M
2=	Past	PAST	147	M
B	budu + Inf.	FUTi	8	P
p=	Perfective Verb	PFVd	104	M

In the sample, there are two West Slavic (Polish and Czech), one East Slavic (Russian) and one South Slavic language (Bulgarian). As pointed out in chapter 3, the Slavic aspect systems have played a very important role in the study of aspect. The status of the Slavic aspectual systems in a general theory of aspect is discussed in detail in that chapter. Here, we shall just note that the Polish, Czech and Russian systems are very similar, whereas Bulgarian differs by having preserved the old Indo-European system alongside the more recently developed Slavic one.

(IQ) Iroquois

Table 6.19 Major TMA categories: Iroquois

Code	Description	Analysis	Frequency	Marking type
Oneida				
22	Perfective Past	FRAMEPAST	8	M
3	Future	FUT	25	M
a	Aorist	PFV	97	M
c	Continuative	<FUT	9	M
p=	Pfv (Stative)	PFCT/STAT	61	M
s	Serial	IPFV	37	M
Seneca				
D	Descriptive aspect	DEFAULT	79	M
I	Iterative aspect	HAB	18	M
Pf	Punctual future	FUT	32	M
Pi	Punctual indicative	PFV	113	M
R	Repetitive particle	HABPAST	13	P
<i>Oneida paradigm: -yoʔt-/yo-t- 'work'</i>				
s	loyó-tas 'he works/is working'			
p=	loyo-té 'he is working'			
22	loyo-tékhwe? 'he worked'			
a	wahoyo-tá 'he worked'			
3	ahoyo-tá 'he will work'			

Iroquois is represented by Oneida and Seneca. Both these are textbook examples of typical American Indian polysynthetic languages. The TMA systems are rather similar to each other, although there are several unclear points. Thus, the category called 'serial aspect' in extant descriptions of Oneida (e.g. Lounsbury 1952) rather seems to be used as a general IPFV in our material. It is not clear if this is due to a dialect difference or what.

(KT) Kam-Tai

Table 6.20 Major TMA categories: Kam-Tai

Code	Description	Analysis	Frequency	Marking type
Thai (Siamese)				
C=	Aux. cá + V	FUT	26	P
D=	Aux. daây + V	?	8	P

K=	Aux. <i>kamlag</i> + V	PROG	18	P
Kh	Aux. <i>khəəy</i> + V	EXPER	7	P
L	V + Aux. <i>læəo</i>	PFCT	26	P
U	Unmarked	DEFAULT	246	U
Y	V + Aux. <i>yuú</i>	PROG	20	P

From this group, we have Thai as a representative: it is an isolating language with a number of periphrastic constructions – a rather typical TMA system for the area.

(MK) Mon-Khmer

Table 6.21 Major TMA categories: Mon-Khmer

Code	Description	Analysis	Frequency	Marking type
Kammu				
C	<i>cəə</i> + V	FUT	24	P
H	<i>hóc</i> + V	PFCT	24	P
K	<i>kù</i> + V	HAB	9	P
U	Unmarked	DEFAULT	184	U

Kammu, which is a language spoken in Northern Laos, Vietnam, Thailand and Southern China, has a TMA system rather similar to that of Thai.

Sketch: Kammu.

(NC) Niger-Congo

(NC1) Bantu

Table 6.22 Major TMA categories: Bantu

Code	Description	Analysis	Frequency	Marking type
Kikuyu				
ag=	Suffix <i>-ag-</i>	IPFV	11	M
ir=	Suffix <i>-ir-</i>	<PFV	13	M
it=	Suffix <i>-it-</i>	<PFCT	16	M
ki=	Prefix <i>ki-</i> : suffix 0	NARR	17	M
ku=	Prefix <i>ku-</i>	HOD FUT?	7	M
ra=	Prefix <i>ra-</i> : suffix 0	PRES-PROG	43	M
rä=	Prefix <i>ra-</i>	HEST	48	M
ä=	Prefix <i>a-</i>	REM-PAST	51	M

Table 6.22 continued

ä=	Prefix <i>a-</i> : suffix 0	IMMED-PAST/PFV?	23	M
Sotho				
11	Simple Present	DEFAULT	59	M
21	Narrative Past	NARR	27	M
23	Recent Past (Perfect)	PFCT/HOD-P	36	M
Ea	Aux. <i>ea</i> + Dep. form	PREHOD-PFV	31	P
Ka	Aux. <i>ka</i> + Dep. form	EXPER	9	P
Ne	Aux. <i>ne</i> + Dep. form	PASTi	37	P
Nt	Aux. <i>ntse</i>	PROG	12	P
Tl	Aux. <i>tla</i>	FUT	17	P
TlBe	<i>tla be</i> + Dep. form	FUT IPFV	6	P
Zulu				
1=	Present	DEFAULT	43	M
21	Narrative Past	NARR	34	M
22 et al.	Prefix <i>-be-</i>	IPFV	99	M
23	Perfect	PFCT/HOD-PFV	65	M
31 et al.	Prefix <i>-zo-</i>	HOD-FUT	16	M
33 et al.	Prefix <i>-yo-</i>	POSTHOD-FUT	8	M
Ya	23 of Aux <i>-ya-</i> + 21	PREHOD-PFV	16	P

The Bantu languages have as a group the most complex TMA systems in the sample: this is easily illustrated by some simple statistics – the average number of ‘major TMA categories’ in the languages of the sample is 6.5, whereas the corresponding average in the Bantu group is 11. To some extent, the remoteness categories which are characteristic of the Bantu languages (see chapter 4) account for this fact, but it should be emphasized that it is not the case that the interest that Bantu speakers seem to pay to remoteness distinctions diverts them from other TMA-relevant notions: on the contrary, it rather seems to have stimulated the development of other dimensions of the TMA systems. Morphologically, the Bantu languages are peculiar in having a rich system of prefixes. The most famous among these are probably the class markers, which appear on various word classes, on verbs as subject and object markers, but also TMA categories (in particular tense categories) and e.g. negation are expressed by prefixes. One might think that the rich morphological systems of the Bantu languages would obviate the need for periphrastic constructions: again, there is no trading relation here – the TMA systems of the

Bantu languages abound in categories expressed this way. (The distinction between what counts as a prefix and what is an auxiliary is often mainly one of orthography.) This is linked up with a tendency to use auxiliaries or 'tense prefixes' also for categories which are in other languages more often expressed by adverbs – there are e.g. forms called 'still-tenses', 'yet-tenses' etc. in Bantu grammars. One peculiarity found in Sotho and Zulu are the so-called 'now-tenses' marked by the prefix/auxiliary *se*, which in the questionnaire is used for events that take place unexpectedly or suddenly, e.g. in (Q.153) (as a translation of English *already*) or in (Q.167):

(Q.153:SO) Morena o se a fihlile
king SM SM arrived
'The king has arrived (already)'

(Q.167:SO) Ka be se ke hata noha
I I step-on snake
'Suddenly I stepped on a snake'

The abundance of TMA categories makes it somewhat difficult to fit the Bantu languages into a universal scheme. The different dimensions of the systems interact in rather complex ways, and the large number of possibilities results in small numbers of examples for each category. Rather a lot of the analyses remain problematic, therefore. We may note some peculiarities: Sotho and Zulu, the two languages from the Nyuni subgroup, apparently both have a PFV/IPFV distinction in the future, something that is rather uncommon (see chapter 3). Typical of the Bantu languages are also the special narrative tenses.

(NC2) Gur

Table 6.23 Major TMA categories: Gur

Code	Description	Analysis	Frequency	Marking type
Karaboro				
B	<i>bà</i> + V	PASTs	48	P
C	<i>má</i> + V	PRES-PROG	10	P
D	dependency marker <i>ń</i>	NARR	27	P
G	<i>gā</i> + V	FUT	26	P
N	<i>nāa</i> + V	PASTi	17	P
U	Unmarked	DEFAULT	69	U
Y	Particle <i>yaà</i>	ALREADY	7	P
h	High tone on subject	IPFV	24	M

Karaboro This language, which belongs to the Gur family and is spoken in Upper Volta, has a rather complicated and interesting TMA system, which it is worthwhile to discuss in some detail. *Karaboro* marks TMA categories by several means: particles, tonal changes in the subject pronoun and the verb, and suppletive series of pronouns. There are two subsystems, the 'dynamic' and the 'stative'. In the latter, present time reference is unmarked, as in:

(Q.1:KB) kē káʔa gbōʔō 'The house is big'
the house big

(Q.30:KB) kē lōʔō yīn 'The water is cold'
the water cold

Past time reference is marked by adding the particle *ba*:

(Q.3:KB) kē káʔa bá gbōʔō 'The house was big'
(Q.32:KB) kē bá yīn 'It (the water) was cold'

Future time reference is marked by the particle *gā*:

(Q.36:KB).lōʔō gā yīn yá gāa 'The water will be cold then'

In the dynamic subsystem, an unmarked verb normally has past time reference and is understood as being perfective, e.g.:

(Q.56:KB) ù kū 'He died/has died'
(Q.46:KB) ñ ũ wāa ʔà 'I met him'
I him meet

The marker *ba* can be added here too. It is used in cases where English has a pluperfect:

(Q.67:KB) yīŋfūa bá kù 'The king has died/was dead'

but also in some other contexts, possibly as a marker of a 'temporal frame' in the past, cf.:

(Q.128:KB) 'yoo-káʔa nge fāa (Looking at a house:) 'Who built this house?'
(Q.129:KB) 'yoo-bà káʔa nge fāa (Looking at a picture of a house which has been torn down:) 'Who built this house?'

For a more problematic example, cf.:

(Q.13:KB) ù bà lētēt yrūhū 'He wrote letters (after dinner)'

In narrative contexts, a so-called 'dependency marker' *ń* (subject to various sandhi changes) is added before the otherwise unmarked verb. Thus, the last sentence of most of our narrative texts reads in *Karaboro* as follows:

(Q.165:KB) ké ŋ kū 'It died'

Future time reference in the dynamic subsystem is marked in the same way as in the stative, by the particle *gā*.

We are now getting to the trickier parts of the system. By changing the tone of the subject to a high one, a construction is obtained which can be used in generic and habitual sentences, e.g.:

(Q.18:KB) ú lētēr yrūhū 'He writes letters'

By adding another subject pronoun and a copula to this sentence, we obtain a construction which is used as a present progressive:

(Q.5:KB) ú mà ú lētēr yrūhū 'He is writing letters'

The following sentence, where the marker *nāa* is used (with what apparently is level tone on the subject), is the counterpart of both these with past reference:

(Q.9 and Q.20:KB) ù nāa lētēr yrūhū 'He was writing/used to write letters'

It appears to me that the simplest way of sorting out these rather complex facts is to assume that the high tone on the subject in (Q.18:KB) is a marker of IPFV which may be neutralized when some other marker appears, e.g. the PASTi marker *nāa*. *Ba*, on the other hand, is what we have above called a PASTs, marking past time reference in stative and pluperfect.

It should also be pointed out that it is possible to transfer stative verbs to the dynamic subsystem: they then obtain an inchoative interpretation.

(NC3) *Kwa*

Table 6.24 Major TMA categories: *Kwa*

Code	Description	Analysis	Frequency	Marking type
Isekiri				
K	<i>ká</i> + V	HABG	22	P
R	V + <i>rē</i>	PFCT	18	P
U	Unmarked	DEFAULT	128	U
W	<i>waá</i> + V	FUT	36	P
WiG	<i>wínɔɔ gbá</i> + V	PROG	16	P
Yoruba				
M=	<i>ma</i> + V	<FUT	6	P
Mpr	<i>ma</i> + <i>n</i> + R	HAB	6	P
U	Unmarked	DEFAULT	144	U
Y	<i>yio</i> + V	FUT	8	P

Table 6.24 continued

p=	<i>ti</i> + V	PFCT	26	P
pr	<i>n</i> + V	PROG	21	P

The two languages from this group – Isekiri and Yoruba – are characterized by periphrastic TMA constructions.

Sketch: Yoruba.

(NC4-5) *Volta-Comoe and West Atlantic*

Table 6.25 Major TMA categories: *Volta-Comoe*

Code	Description	Analysis	Frequency	Marking type
Akan				
1=	Present	DEFAULT	42	U
2	Past	PFV	96	M
3=	Future	FUT	16	M
NA	<i>na</i> + V	PAST?	41	P
T=	<i>-taa</i> + V	HAB	8	P
p=	Perfect	PFCT	22	M
pr	Progressive	PROG	23	P

Table 6.26 Major TMA categories: *West Atlantic*

Code	Description	Analysis	Frequency	Marking type
Wolof				
20	Past	PAST	71	M
20Na	Past + <i>na</i>	N-NARR		
		PFV/PLPFCT	35	P
FD21	<i>dan</i> + <i>de</i> + V	HABPAST	7	P
F	Aux. <i>dafa</i> + V	STAT/FUT?	59	P
FD	Aux. <i>dafa</i> + <i>de</i> + V	HABG	18	P
M=	Aux. <i>munge</i> + V	PROG?	9	P
Na=	V + <i>na</i>	PFCT	29	P
U	Unmarked	DEFAULT	65	U
Y=	<i>dina</i> + V	FUT	16	P
FD20	<i>dafa</i> + <i>don</i> + V	PAST-PROG	13	P

From these two groups, there are two languages with rather complex TMA systems, Akan and Wolof, which are notable above all for the

complex ways in which past time reference is signalled (see chapter 4 for discussion).

Sketches: Akan, Wolof.

(ST) Sino-Tibetan

Table 6.27 Major TMA categories: Sino-Tibetan

Code	Description	Analysis	Frequency	Marking type
Chinese (Mandarin)				
G	Particle <i>guo</i>	EXPER	23	P
H=	Particle <i>hui</i>	<FUT	11	P
L	Particle <i>le</i>	PFV	64	P
U	Unmarked	DEFAULT	117	U
Z=	Particle <i>zài</i>	PROG	17	P
ZH	Particle <i>zhe</i>	<PROG?	7	P

This group is represented in the sample only by Mandarin Chinese (Putonghua). There is an extensive literature about the TMA categories of this language, to which I will not have too much new to add. On the basis of the questionnaire material, it is possible to identify the categories PFV, EXPER and PROG. It should be noted, however, that this is most probably an incomplete list: the particle *le*, which we have labelled PFV, apparently has two functions which correspond to two possible positions: directly after the verb it is indeed a PFV marker, but when it appears in sentence final position it is apparently used more like a PFCT (or possibly a RESUL) – see e.g. Li et al. 1982 for a discussion. However, in many contexts, the PFV *le* and the sentence-final *le* will be homonymous, e.g. in the following:

(Q.67:CH) guówáng shì le ‘The king died/has died’

It is therefore not possible to separate the two on the basis of our material.

(UC) Uralic

Table 6.28 Major TMA categories: Finno-Ugric

Code	Description	Analysis	Frequency	Marking type
Finnish				
1=	Present	DEFAULT	43	M
2=	Simple Past	PAST	134	M

Table 6.28 continued

O	Aux. <i>olla</i> + Supine	PFCT	62	P
OX	<i>on</i> + 3rd Inf. in inessive	<PROG	7	P
O2	Past of Aux. <i>olla</i> + Supine	<PLPFCT	18	P
Hungarian				
1	Present	DEFAULT	130	M
2	Past	PAST	154	M
3	Future of Copula	FUTs	6	M
F	Aux. <i>fog-</i> + Inf.	FUT	12	P
p=	Perfectivizing prefixes	PFVd?	70	M
Estonian				
10	Present	DEFAULT	116	M
20	Past	PAST	144	M
O	Aux. <i>ola</i> + PtP	PFCT	49	P
i	Modus Obliquus (Indirect Speech Form)	<QUOT	9	M
O2	Past of Aux. <i>ola</i> + PtP	PLPFCT	14	P

The Uralic family is represented by the closely related Finnish and Estonian and by Hungarian. These languages are notable above all for the ways in which perfectivity interacts with transitivity: in Finnish and Estonian, mainly in the case marking system, and in Hungarian, in the ‘perfectivizing’ verbal prefixes or particles (see chapter 3 for discussion).

Sketch: Hungarian.

7

Conclusion

The major claim made in this book has been that it is possible to reduce the bewildering multitude of tense-mood-aspect categories found in the languages of the world – or at least the overwhelming majority of such categories – to a fairly small set of cross-linguistic categories, characterized by bundles of morphosyntactic and semantic properties, more specifically by the (proto)typical contexts in which they are used and by the typical ways in which they are manifested morphologically or syntactically. I have tried to substantiate this claim by playing a certain kind of ‘analysis game’ on a data base. It is now up to the reader to judge the success of the program. In my opinion, the data fit the hypothesis well enough – given the circumstances – for us to consider it empirically supported. Let us try to sum up the facts. Among the categories tentatively identified in the coding and analysis stages of the primary investigation, a subset were selected according to an operational definition of ‘major TMA category’. In table 7.1, these ‘major TMA categories’ are broken down as to the analyses that were subsequently assigned to them. It can be seen that in the overwhelming majority of cases, it has been possible to assign those language-specific categories to one of a relatively small set of hypothesized cross-linguistic categories. In most cases, the use of a cross-linguistic label for a language-specific category means that the distribution of the category has been found sufficiently similar to the predicted distribution of the cross-linguistic category. In other words, the fit between the predicted and the actual distribution in the questionnaire has been shown to be as good as can be expected under the circumstances. On the negative side, there are still many unclear points, and many of the categories behave in ways which we still do not understand. However, I would personally feel less well at ease if these problems did not exist – too perfect a body of data reminds one of the election results of some countries.

Whether the reader feels convinced will partly depend on whether I have managed to present the material in a clear enough way. I have tried to put the cards on the table as openly as possible, but it is

Table 7.1 Major TMA categories in the material by cross-linguistic category type and marking type

Cross-linguistic category assigned	Unqualified analysis				Analysis qualified by ‘<’ or ‘?’		
	Marking type			Total	Marking type		
	M	P	U		M	P	Total
ALREADY		1		1			
CONCL		1		1			
DEFAULT	37		26	63		2	2
DEFAULTs		1	2	3			
EXPER		7		7	1		1
FRAMEPAST	1			1			
FUT	25	19		44	2	3	5
FUT/IPFV	1			1			
FUTi		5		5			
FUTs	1			1			
HAB	2	3		5			
HABG		3		3	1		1
HABG/PRED	1			1			
HABPAST	1	2		3			
HABPASTc	1			1			
HEST	1			1			
HOD-FUT				1		1	
IMMED-PAST		2		2			
IMMED-PAST/PFV				1		1	
IPFV	7			7	1	1	2
IPFV/RESUL		1		1	1		1
NARR	3	1		4			
N-NARRPFV/PLPF		1		1			
PAST	15	2		17		1	1
PASTi	13	8		21			
PASTid	1			1			
PASTn	2	1		3		1	1
PASTs	2	1		3		4	4
PAST-PROG		1		1			
PFCT	2	16		18	2	4	6
PFCTq	1	1		2			
PFCT/STAT	1			1			
PFCT/HOD-PFV	2			2			
PFV	37	6		43	3		3
PFVd	4			4			
PFV N-VOL	1			1			
PFV VOL	1			1			
PLPFCT	2	16		18		2	2
POSTHOD-FUT	1			1			

Table 7.1 continued

PRED	1	1	2	1	1	
PREHOD-PFV		2	2			
PRES-PROG	2	3	5			
PROG	1	18	19	1	6	7
PROG/STAT	1		1			
PROSP	1		1			
QUOT	3	1	4	1		1
REM-PAST	1		1			
STAT/FUT					1	1
?		1	1	2		

difficult to do justice to data that are as complex as the ones we are dealing with here.

It should be emphasized that the data I have had at my disposal for each language have been quite limited, and I suspect that many of the analyses I have made of language-specific categories may be incorrect. It seems to me, however, that what we need at this point is a rough mapping of the hitherto relatively unknown territory in order to be able to go further. What I have wanted to give is a set of falsifiable claims that can be a point of departure for future studies.

Let us now look at the general picture of what tense-aspect systems are like that emerges from our investigation. In addition to the finding that only a relatively restricted set of category types tend to be involved in tense-aspect systems, what is most important to emphasize is that these category types differ as to the role their representatives typically play in the systems. There is in my opinion good support for the view that tense-aspect systems are organized in terms of a 'centre' and a 'periphery', where the set of central categories is even more restricted than that of tense-aspect categories in general. The centrality of a category shows up in several ways. Most importantly, one of the more striking findings of this study is the high correlation between our hypothesized cross-linguistic TMA categories and types of marking – morphological or periphrastic. This issue demands to be treated at some length.

The distribution of morphological and periphrastic categories is summed up in table 7.1.¹ It turns out that if we know what cross-linguistic TMA category a language-specific category belongs to, we are not only able to predict its distribution in the questionnaire with a precision of up to 80 per cent but we can also in some cases predict whether it is expressed morphologically or periphrastically with at least the same accuracy. For example, out of 17 clear instances of the (unrestricted) category PAST, 15 are marked morphologically,

whereas out of 19 clear instances of the category PROG, only 1 is. Another way of putting it is to observe that out of 155 morphologically expressed categories, 117, i.e. roughly 75 per cent, belong to a group of three, viz. PAST, PFV:IPFV and FUT (varieties included). This is further illustrated in table 7.2, which lists the morphologically expressed categories of each language in the sample. It should be emphasized that criteria of form played no role whatsoever in the assignment of cross-linguistic labels to language-specific categories: there would thus be no a priori reason to expect that they would differ in type of marking. This makes it plausible to assume – as was suggested in chapter 1 – that there is indeed a close connection between the semantics of a TMA category and the ways in which it tends to be expressed. The idea – which will certainly have to be made more specific – is that only categories with a 'Boolean' semantics (as this notion was explicated in chapter 1) will be frequently expressed by inflectional categories. The obvious reason is that it is only 'Boolean' categories for which the restricted expressive power of inflectional processes is sufficient. Inflectional categories do not in general allow for iteration or alternative orders of application, phenomena that are essential for categories with an 'operator' logic.

We have so far been talking of a simple dichotomy between morphologically and periphrastically expressed categories, identifying 'morphological' with 'inflectional'. In the chapter on aspectual categories, one of the main points made was that what we have called PFV:IPFV categories differ as to their position on a scale from typical 'inflectional' to typical 'derivational' categories and that this is correlated with differences in a number of semantic parameters. We thus have a further illustration of the interplay between form and content in tense-aspect categories that we talked about above.

Among inflectional categories, there is of course great variation as to the ways in which they are marked. Since a reliable classification demands relatively subtle grammatical analyses of the individual languages, I cannot offer any statistics here, but it appears that the general tendencies agree with what has been established in earlier studies concerning morphological categories in general. Among other things, suffixation appears to be much more common than prefixation. Furthermore, it is a relatively restricted group of languages that exploit prefixation systematically for the marking of TMA categories. The most notable group here are the Bantu languages. It is striking that the Bantu languages are also the group that have the most complex TMA systems in general. These facts invite some speculations. A classical problem in Bantu grammar – and in Bantu orthography – is the determination of the borderline between bound and free grammatical morphemes. In other words, it is often an open question whether a certain TMA marker should be regarded as a prefix or rather as e.g. an auxiliary. It was also noted that the Bantu

Table 7.2 Major TMA categories with morphological expression

Cushitic	Beja	⊂ PFV/IPFV PAST _i d
	Oromo (Galla)	⊂ <PFV/IPFV PFCT
Semitic	Arabic (Modern Standard)	⊂ PFV/IPFV FUT
	Arabic (Tunisian)	⊂ PFV/IPFV
	Hebrew	⊂ PAST FUT
	Maltese	⊂ PFV/IPFV
	Tigrinya	⊂ PAST FUT PRES-PROG
	Amharic	⊂ PAST PFCT
Turkic	Azerbaijani	⊂ PFV/IPFV PFCT _q PLPFCT FUT PRED PAST _i
	Turkish	⊂ PFV/IPFV HABG/PRED FUT PAST _i QUOT
Korean-Japanese	Japanese	⊂ <PFV/IPFV
Andean-		
Quechumaran	Quechua	(⊂) PAST _n FUT QUOT PROG/STAT
Tupi	Guarani	(⊂) PAST _n FUT <PROG
AUSTRALIAN	Alawa	⊂ PFV/IPFV PAST _i HABPAST
	Bandjalang	⊂ PFV/IPFV PAST _i FUT PROG
West Indonesian	Cebuano	⊂ PFV/IPFV <PRED <EXPER HABG
	Indonesian	⊂ —
	Javanese	⊂ —
	Sundanese	⊂ —
South Celebes	Bugis Makassar	⊂ —
Polynesian	Hawaiian	⊂ —
	Maori	⊂ PFV/IPFV?
Kartvelian	Georgian	⊂ PFV/IPFV PAST _i FUT PFCT _q
DRAVIDIAN	Tamil	⊂ PAST FUT/HAB?
ESKIMO-ALEUT	Greenlandic Eskimo	⊂ PFCT FUT QUOT CONCL? HABG
Germanic	Afrikaans	⊂ PAST _s
	English	⊂ PAST
	Fitzroy Crossing	⊂ —
	Kriol	⊂ PROG
	German	⊂ PAST
	Swedish	⊂ PAST
Indo-Iranian	Bengali	⊂ PAST HABPAST _c FUT
	Hindi/Urdu	⊂ PAST _i PAST _s
	Kurdish	⊂ PFV/IPFV PAST _i FUT
	Punjabi	⊂ PFV/IPFV PAST _s
	Persian	⊂ PFV/IPFV PAST _i
Greek	Greek (Modern)	⊂ PFV/IPFV PAST
Romance	Catalan	⊂ PAST _i <NARR FUT
	French	⊂ PFV/IPFV PAST _i FUT
	Italian	⊂ PFV/IPFV FUT PAST _i
	Latin	⊂ PFV/IPFV PAST _i PLPFCT FUT
	Limouzi	⊂ PFV/IPFV PAST _i FUT

Table 7.2 continued

	Romanian	<PFV/IPFV PAST _i
	Spanish	⊂ PFV/IPFV PAST _i FUT
	Portuguese	⊂ PFV/IPFV PAST _i <FUT
Slavic	Bulgarian	⊂ PFV/IPFV PAST _s /i PFV _d QUOT
	Czech	⊂ PFV _d
	Polish	⊂ PAST PFV _d
	Russian	⊂ PAST PFV _d
IROQUOIS	Oneida	FRAMEPAST FUT PFV/IPFV PFCT/STAT
	Seneca	HAB FUT PFV/IPFV HABPAST
KAM-TAI	Thai (Siamese)	—
MON-KHMER	Kammu	—
Bantu	Kikuyu	⊂ PFV/IPFV <PFCT NARR HOD-FUT PRES-PROG HEST REM-PAST IMMED-PAST/PFV
	Sotho	NARR PFCT/HOD-PFV
	Zulu	NARR IPFV PFCT/HOD-PFV HOD-FUT POSTHOD-FUT
Gur	Karaboro	⊂ PFV/IPFV
Kwa	Isekiri	—
	Yoruba	—
Volta-Comoe	Akan	⊂ PFV/IPFV FUT PFCT
West Atlantic	Wolof	⊂ PAST
SINO-TIBETAN	Chinese (Mandarin)	—
Finno-Ugric	Finnish	⊂ PAST
	Hungarian	⊂ PAST FUT _s PFV _d ?
	Estonian	⊂ PAST <QUOT

languages abound in periphrastic constructions which play a more or less peripheral role in the TMA systems. One might thus speculate that the use of prefixal categories in a language – in itself a ‘marked choice’ – makes the system of inflectional categories more open in the sense that the borderline to syntactically marked categories is less well defined. A consequence of this might then be that a prefixing language would in general allow more differentiation in what categories are morphologically expressed.

Something that might be worth exploring further is the connection between the general morphological type of a language and what TMA categories it has. Given that e.g. PAST is very infrequently expressed by any other means than inflection, a language which does not in general have inflection will not be likely to have PAST as a major TMA category. In general, the languages in our sample appear to behave in accordance with this prediction.

In addition to marking type, there are a few other plausible criteria

for determining the centrality of a tense-aspect category. One which we have talked about earlier (p.23) is obligatoriness of expression. This is something which it is much harder to make judgements about, and no very clear tendencies appear. Still, the general feeling is that the categories PAST and PFV:IPFV tend to be obligatory when they exist in a language, whereas many of the categories assumed to be less central include several clear optional cases.

The question of obligatoriness is linked up with the issues discussed in the section 'Accidence categories and Gricean principles' in chapter 1. One observation we may make in this connection is that if we compare the cross-linguistic categories PAST and PFCT, of which the former is commonly expressed morphologically and the latter periphrastically, there is a rather clear difference in their 'redundancy' in the sense that PFCT is much less often used in contexts where the time reference is already given. In other words, PAST is a much clearer example of an 'accidence category' in this regard.

An easier criterion to apply is that of frequency. For each cross-linguistic category, we may talk of at least two frequency measures: (i) the number of languages in which the category is instantiated; (ii) the text frequency of each language-specific instance of the cross-linguistic category. As to the first criterion, we may note that the three categories FUT, PAST and PFV:IPFV are the most commonly represented categories in table 7.2. In other words, this criterion singles out the same group as that of type of marking: the categories that are most often expressed morphologically are also the most common in general. Concerning the second criterion, that of text frequency, it is harder to make judgements, since we can only talk of frequencies in the questionnaire, which may not at all correspond to frequencies in spontaneously produced corpora. Still we may note that two categories – PAST and PFV:IPFV (or rather the marked member of the latter in each language) – are outstanding in having much higher average frequencies than any other categories in the questionnaire. Summing up, we may conclude that the different criteria of centrality assign roughly the same order to the cross-linguistic categories we have postulated, with PAST, PFV:IPFV and FUT as a clear central group.

A further criterion, which we have discussed in passing in earlier chapters, is that of 'relative precedence', i.e. which one out of two categories will be used if the conditions for both are fulfilled. Here, it appears relatively clear, at least for a number of cases, that the group we have called 'central' are also the losers in conflict cases – that is, they are chosen only if the conditions for the other category do not apply. One example might be a language (such as English) which has both PFCT and PAST – effectively, the distribution of the former is included in that of the latter, which means that PFCT has to win the conflict in order to surface at all. Many linguists would describe such a

relation between categories in terms of markedness – in general, it is clear that the concept of centrality that I am discussing here is closely related to what has been studied under that heading: the more peripheral a category is, the more 'marked' (in the wide sense of 'markedness') it is.

Let us briefly survey the hypothesized cross-linguistic categories that are assumed to build up TMA systems in human languages.

Among the typical aspectual categories, we find the PFV:IPFV opposition and the PROGRESSIVE. Whereas the latter tends to be periphrastic and is in most cases independent of time reference, the former is typically expressed morphologically and is closely linked up with the tense category PAST in at least two respects: (i) PFV is normally used only with past time reference, (ii) PAST is applied only to IPFV contexts. When both these conditions are fulfilled, the two categories make up a tripartite system, as exemplified e.g. in many Indo-European languages, which can frequently be seen as the core of the TMA system. There are deviant cases, however, the most notable ones being found in the Slavic languages, where PFV:IPFV and PAST operate wholly independently of each other. As noted above, this can be related to the fact that the Slavic PFV:IPFV is realized as a derivational rather than as an inflectional category.

It should also be noted that it is not uncommon for a language to lack one or both of the two categories we have discussed.

PAST takes us to the categories traditionally regarded as tenses. Among those, one category that comes close to being universally represented is FUTURE, which, however, scores lower on some other parameters of centrality, in being more often expressed periphrastically than e.g. PAST and being less often obligatory. We also noted the existence of more marginal future signalling categories, such as PREDICTIVE and PROSPECTIVE. Among tense categories a special place is taken by the categories that express remoteness distinctions. In this area, we noted a striking uniformity in the choice of 'cut-off points' and also some very widespread tendencies for the realization of remoteness nuances as secondary meanings of other TMA categories, such as PERFECT and PLUPERFECT.

We now move to the less central categories of tense-aspect systems: mostly, these are traditionally regarded as 'aspects'. One such group is that of habitual categories, HABITUAL, HABITUAL-GENERIC and HABITUAL-PAST. One result of our investigation is the finding that although the notions of habituality and genericity play an important role in many TMA systems, it is less common for them to be represented by independent categories: in particular, it was not possible to identify any case of a marked category restricted to generic contexts only. Something similar may be said of the concept of narrativity, which is often important for the choice between TMA categories, but which is seldom marked independently – in this case

we cannot turn 'seldom' into 'never' however: there are several examples of what we have called NARR in the material.

One important family of categories is that which includes the PERFECT, the RESULTATIVE, the EXPERIENTIAL and the PLUPERFECT. The semantic relations between these categories turned out to be significantly more complex than had been thought: neither the RESULTATIVE nor the EXPERIENTIAL can be regarded simply as more restricted cases of the PERFECT. On the other hand, the latter was established as a relatively uniform category cross-linguistically.

Finally, we should mention QUOTATIVE, which is the most clearly modal category among the 'major TMA categories' we have been looking more closely at.

There are many approaches to the study of TMA categories: one may choose among a bewildering multitude of theoretical frameworks, and among many different kinds of data. In this book, the approach has been typological. We have tried to draw conclusions about cross-linguistic generalizations from questionnaire data from a wide range of languages. This has clearly biased the ways in which different aspects of TMA systems have been treated: for instance, I have not been able to pay as much attention to things like language acquisition and language change as I have would liked to, but I hope that this will not be taken to mean that I regard those aspects as unimportant. On the contrary, I think that the explanations of many of the descriptive generalizations that have been made in this book will only be found when we look closer at how people learn to use TMA systems and how these systems change. In general, people may find that I have been rather cautious in offering explanations for my findings. It should be clear from the discussion of universals in chapter 1 that I find explanations in terms of a static 'universal grammar' rather unfruitful. My feeling is that such universals as there are have to be explained by an interplay of many factors, and that the importance of natural languages' character of 'self-organizing' social systems for information processing has been underrated in this connection.

Notes

Chapter 1 General background

- 1 I use the term 'gender' as a cover term both for the gender category in e.g. Indo-European and for similar phenomena such as the noun classes of the Bantu languages.
- 2 Cf. Thelin (forthcoming), where it is said e.g. 'Proceeding from the same principles according to which we can say that all languages have aspect, we can also say that they have at their common disposal basic temporal oppositions. Slavic languages with one single tense form for past time as, for example, Russian do not lack those temporal meanings of past time (nor, actually, corresponding means of expression) which Bulgarian expresses formally and systematically.' It may be noted that the example Thelin goes on to quote, i.e. the purported covert category of pluperfect in Russian, is somewhat atypical of TMA categories in that the distinction between Simple Past and Pluperfect in e.g. English quite often corresponds to a very clear distinction in truth-conditions. Cf.:
 - (i) Yesterday I finished two chapters.
 - (ii) Yesterday I had finished two chapters.
 Correspondingly, sentences like (ii) belong to the quite restricted number of cases where the lack of a TMA category in a language leads to the need to compensate the loss in information by other means, for instance, as is often done in Russian, by adding a word like *uže* 'already'.
- 3 The problems of trying to do semantics with binary features or equivalent constructs were made clear in the early discussion of Katz's semantic theory (see e.g. Weinreich 1966, Bierwisch 1969).
- 4 Cf. e.g. his discussion of the impossibility of giving a universal definition of a term like 'the subjunctive mood' (1924, 48).
- 5 As noted in that paper, the terminology is extremely confusing. Basically, the distinction made here will correspond to Vendler's distinction between 'accomplishments/achievements' on the one hand and 'activities' on the other, and to what some authors call 'telic/atelic'. The term 'bounded' ('having a bound or limit') is sometimes misunderstood as 'bound' (past participle of 'bind'): one Russian aspectologist once quoted me as having introduced the concept 'svjazannyj' into aspectology.

Chapter 2 The investigation

- 1 Here and below we sometimes use 'verb' – sloppily but conveniently – to refer to nominal verb-less predicate constructions, which occur in a number of languages.
- 2 The following comment on (Q.56) by A.P. Omamor, our Isekiri consultant, is worth quoting in this connection: 'The verb *kú* "die" is never used to refer to the king, it is taboo; the king does not die, he joins his ancestors, as it were.'
- 3 System 1022 is a general purpose data management software system which operates on DEC10 and DEC20 computers. It is sold by Software House, 1105 Massachusetts Avenue, Cambridge, MA 02138, USA.

Chapter 3 Aspectual categories

- 1 The above observations on the interaction of aspect and definiteness in Hungarian are due to Dora Kós-Dienes.

Chapter 4 Tense categories

- 1 The lower C-value given in table 4.1 was computed for both the Present and the Past of *skall* – the higher one is obtained if we count the Present only.
- 2 Ultan 1978, 94 seems to want to suggest that it is the other way round, although his claim is rather hedged. One of his examples of languages where 'the occurrence of future tenses is obligatory as opposed to optional occurrence of past tenses in certain contexts', viz. Guarani, is directly contradicted by our material. There are several occurrences of a Present used with future time reference in our Guarani questionnaire (e.g. Q.151, 792, 1042).
- 3 Most of the material contained in this section was presented in Dahl 1984a.
- 4 This statement is actually an oversimplification, holding only for past tenses. See Hymes's paper for an account of the superficially contradictory behaviour of the prefixes *t-* and *u-*.

Chapter 5 The Perfect (PFCT) and its relatives

- 1 More precisely, he claims that it will then have a Future Perfect interpretation.
- 2 There is a possible example in our material which is clearly outside the 'Ottoman' area, viz. Greenlandic Eskimo. The QUOT category listed in table 5.11 for this language employs the same suffix, *-sima*, as another category which we have labelled <PFCT. I do not know the reason for not analysing these categories as one and the same – if this is done, the resulting category turns out to pass the C>.40 criterion for PFCTq.

Chapter 6 An overview of the TMA systems of the languages in the sample

- 1 I am indebted to Magdalena Wichser for supplying information on Karaboro.

Chapter 7 Conclusion

- 1 The figures given are those that hold for the large sample. A check of the small sample shows that the ratios between morphologically and periphrastically expressed categories are the same there, which gives reason for some confidence that we are not dealing with accidental coincidences.

References

- Allwood, J. 1976. *Linguistic communication as action and cooperation. A study in pragmatics*. Univ. of Göteborg, Dept. of Linguistics.
- Andersson, S.-G. 1972. *Aktionalität im Deutschen: Eine Untersuchung unter Vergleich mit dem russischen Aspektsystem*. Uppsala: Acta Universitatis Upsaliensis.
- Angenot, J.-P. 1971. *Aspects de la phonétique et de la morphologie de l'ewondo*. Brussels: Print-Express.
- Appleby, L. L. 1961. *A first Luyia grammar*. Nairobi: East African Literature Bureau.
- Bell, A. 1978. Language samples. In Greenberg, J. (ed.), *Universals of human language*. Vol. 1. *Method and theory*. Stanford: Stanford Univ. Press.
- Berlin, B. & Kay, P. 1969. *Basic color terms, their universality and evolution*. Berkeley: Univ. of California Press.
- Bickerton, D. 1981. *Roots of language*. Ann Arbor: Karoma Publishers.
- Bierwisch, M. 1969. Certain problems of semantic representations. *Foundations of Language* 3. 1-36.
- Butterworth, B., Comrie, B., & Dahl, Ö., (eds). 1984. *Explanations of language universals*. The Hague: Mouton.
- Carlson, G. 1977. A unified analysis of the English bare plural. *Linguistics and Philosophy* 1. 413-58.
- Carlson, G. 1979. Generics and atemporal when. *Linguistics and Philosophy* 3. 49-98.
- Carlson, G. 1982. Generic terms and generic sentences. *Journal of Philosophical Logic* 11. 145-81.
- Chia, E. N. 1976. Kom tenses and aspects. Unpublished dissertation. Georgetown University, Washington, DC.
- Chomsky, N. 1981. *Lectures on government and binding*. Dordrecht: Foris Publications.
- Colarusso, J. 1979. Verbs that inflect for kinship: grammatical and cultural analysis. *Papiere zur Linguistik* 20. 37-66.
- Cole, D. T. 1955. *An introduction to Tswana grammar*. London: Longmans.
- Comrie, B. 1976. *Aspect. An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge Univ. Press.
- Comrie, B. Forthcoming. *Tense*. Cambridge: Cambridge Univ. Press.
- Dahl, Ö. 1975. On generics. In Keenan (ed.) 1975.
- Dahl, Ö. 1977. Games and models. In Dahl, Ö. (ed.), *Logic, pragmatics and grammar*. Univ. of Göteborg, Dept. of Linguistics.
- Dahl, Ö. 1979. Case grammar and prototypes. *Prague Bulletin of Mathematical Linguistics* 31.
- Dahl, Ö. 1980a. Russian direct objects and functional case marking principles. In Studia Slavica Gunnaro Jacobsson sexagenario a discipulis oblata. Slavica Gothoburgensia 7. Stockholm: Almqvist & Wiksell.
- Dahl, Ö. 1980b. Tense-mood-aspect progress report. *Gothenburg Papers in Theoretical Linguistics*.
- Dahl, Ö. 1981. On the definition of the telic-atelic (bounded-nonbounded) distinction. In Tedeschi & Zaenen (eds). 1981.
- Dahl, Ö. 1984a. Temporal distance: remoteness distinctions in tense-aspect systems. In Butterworth et al. (eds). 1984.
- Dahl, Ö. 1984b. Sketch of the TMA system of Beja. In Dahl & Kós-Dienes (1984).
- Dahl, Ö., forthcoming. Aspect and case-marking in the Fenno-Ugric languages.
- Dahl, Ö. & Kós-Dienes, D. 1984. *Selected papers from the Tense-Mood-Aspect Project*. Univ. of Stockholm, Dept. of Linguistics.
- Dahl, Ö. & Talmoudi, F. 1979. *Qad and laqad - tense/aspect and pragmatics in Arabic*. In Pettersson, T. (ed.), *Aspectology - Papers from the 5th Nordic Conference of Linguistics*. Univ. of Lund, Dept. of Linguistics.
- Davies, J. 1981. Kobon. *Lingua Descriptive Studies* 3. Amsterdam: North Holland.
- Derbyshire, D. C. 1979. Hixkaryana. *Lingua Descriptive Studies* 1. Amsterdam: North Holland.
- Fillmore, C. & Langendoen, T. (ed). 1971. *Studies in linguistic semantics*. New York: Holt, Rinehart & Winston.
- Friedrich, P. 1974. On aspect theory and Homeric aspect. *International Journal of American Linguistics*. vol 40, No. 4, Part 2, Memoir 28.
- Grice, H. P. 1975. Logic and conversation. In Cole, P. & Morgan, J. L. (eds), *Syntax and semantics 3: Speech acts*. New York: Academic Press.
- Groot, C. de. 1984. Totally affected. Aspect and three-place predicates in Finnish. In de Groot & Tommola (eds) 1984.
- Groot, C. de & Tommola, H. (eds). 1984. *Aspect bound. A voyage into the realm of Germanic, Slavonic and Finno-Ugrian aspectology*. Dordrecht: Foris Publications.
- Haltof, B. 1968. Ein semantisches Modell zur Aspekt determinierung im modernen Russischen. In *Probleme der strukturellen Grammatik und Semantik*. Leipzig: Karl Marx Universität.
- Hoffmann, C. 1963. *A grammar of the Margi language*. Oxford University Press.
- Hopper, P. 1982a. Aspect between discourse and grammar. Introductory essay for Hopper (ed.) 1982b.
- Hopper, P. (ed.). 1982b. *Tense-aspect: Between semantics and pragmatics*. Amsterdam: John Benjamins.
- Hopper, P. & Thompson, S. 1980. Transitivity in grammar and discourse. *Language* 56. 251-99.
- Hyman, L. 1979. Aghem grammatical structure. *Southern California Occasional Papers in Linguistics* 7. Los Angeles: University of Southern California Press.

- Hymes, D. 1975. From space to time in tenses in Kiksht. *International Journal of Applied Linguistics* 41. 313-29.
- Inoue, K. 1975. Studies in the perfect. Unpublished dissertation, Univ. of Michigan.
- Jakobson, R., Fant, G., Halle, M. 1963. *Preliminaries to speech analysis*. Cambridge, Mass.: MIT Press.
- Javanaud, P. 1979. Tense, mood and aspect (mainly aspect) in Limouzi. *Gothenburg Papers in Theoretical Linguistics* 39. Univ. of Göteborg, Dept. of Linguistics.
- Jespersen, O. 1924. *The philosophy of grammar*. London: Allen & Unwin.
- Johnson, B. 1979. *Hebräisches Perfekt und Imperfekt mit vorangehendem we*. Lund: Gleerup.
- Johnson, M. R. 1981. A unified temporal theory of tense and aspect. In Tedeschi & Zaenen (eds) 1981.
- Kamp, J. 1975. Two theories about adjectives. In Keenan (ed.) 1975.
- Katenina, T. E. 1960. *Jazyk chindi*. Moscow: Nauka.
- Keenan, E. L. (ed.). 1975. *Formal semantics of natural language*. Cambridge: Cambridge University Press.
- Keenan, E. L. 1976. Towards a universal definition of 'subject'. In Li, C. N. (ed.), *Subject and topic*. New York: Academic Press.
- Kopchevskaya, M. 1984. Some Bantu tense systems. In Dahl & Kós-Dienes 1984.
- Kučera, H. 1981. *Aspect, markedness and t_0* . In Tedeschi & Zaenen (eds) 1981.
- Lancelot, C. & Arnauld, A. (1660). *Grammaire générale et raisonnée*. Paris: Pierre le Petit.
- Li, C. N., Thompson S. A., & Thompson, R. M. 1982. The discourse motivation for the perfect aspect: The Mandarin Particle *LE*. In Hopper (ed.) 1982b.
- Ljung, M. 1980. Reflections on the English progressive. *Gothenburg Studies in English* 46. Göteborg: Acta Universitatis Gothoburgensis.
- Lounsbury, F. 1952. Oneida verb morphology. *Yale University Publications in Anthropology* No. 48. New Haven: Yale University Press.
- Lundberg, L.-J. 1984. Perfekt? En verbforms grad av tillämplighet i olika kontexter. Term paper, Univ. of Stockholm, Dept. of Linguistics.
- Martin, S. 1975. *A reference grammar of Japanese*. New Haven: Yale Univ. Press.
- McCawley, J. 1971. Tense and time reference in English. In Fillmore & Langendoen 1971.
- McCoard, R. 1976. *The English perfect: Tense choice and pragmatic inferences*. Amsterdam: North Holland.
- Morolong, M. 1978. Tense and aspect in Sesotho. Unpublished dissertation, Simon Fraser University, Burnaby, British Columbia.
- Mønnesland, S. 1984. The Slavonic frequentative habitual. In de Groot & Tammola (eds) 1984.
- Nedjalkov et al. 1983. *Rezultativnye konstrukcii*. Moscow: Nauka.
- Pettersson, T. 1972. *On Russian predicates: a theory of case and aspect*. Stockholm: Almqvist & Wiksell.
- Quine, W. 1960. *Word and object*. Cambridge, Mass.: MIT Press.

- Rafferty, E. 1982. Aspect in conversational Indonesian. In Hopper (ed.) 1982b.
- Reichenbach, H. 1947. *Elements of symbolic Logic*. New York: Macmillan.
- Russkaja grammatika. 1980. Tom 1. *Fonetika, fonologija, udarenie, intonacija, slovoobrazovanie, morfologija*. Moscow: Nauka.
- Siegel, S. 1956. *Non-parametric statistics: For the behavioral sciences*. New York: McGraw-Hill.
- Slobin, D. & Aksu, A. 1982. Tense, aspect and modality in the use of the Turkish evidential. In Hopper (ed.) 1982b.
- Stojanov, S. 1964. *Gramatika na bulgarskija knižoven ezik*. Sofija: Nauka i izkustvo.
- Tedeschi, P. & Zaenen, A. (eds). 1981. *Syntax and semantics 14: Tense and aspect*. New York: Academic Press.
- Thelin, N. 1978. Towards a theory of aspect, tense and actionality in Slavic. Uppsala: Acta Universitatis Upsaliensis, Studia Slavica 18.
- Thelin, N. Forthcoming. Linguistic typology, universal grammar and the Slavic verb. Forthcoming in *Slavica Gothoburgensia* 10.
- Ultan, R. 1978. The nature of future tenses. In Greenberg, J. (ed.), *Universals of human language*. Vol. 3. *Word structure*. Stanford: Stanford University Press.
- Vendler, Z. 1967. *Linguistics and philosophy*. Ithaca, New York: Cornell University Press.
- Vinogradov, V. V. 1947. *Russkij jazyk*. Moscow: Učpedgiz.
- Weinreich, U. 1966. Explorations in semantic theory. In Sebeok, T. (ed.), *Current trends in linguistics*. Vol. III. The Hague: Mouton.
- Welmers, W. 1972. *African language structures*. Berkeley: University of California Press.
- Whiteley, W. H. & Muli, M. G. 1962. *Practical introduction to Kamba*. Nairobi and London: Oxford University Press.
- Wolff, J. U. 1966. *Beginning Cebuano*. Part I & II. New Haven & London: Yale University Press.

Appendix

The TMA questionnaire

Context indications are given within square brackets. Words within parentheses are not to be translated.

Part A – sentences

- (1) [Standing in front of a house] The house BE BIG
- (2) [Talking about the house in which the speaker lives (the house is out of sight)] The house BE BIG
- (3) [Talking about a house in which the speaker used to live but which has now been torn down] The house BE BIG
- (4) [Talking about a house which the speaker saw for the first time yesterday and doesn't see now:] The house BE BIG
- (5) [Q: What your brother DO right now? (=What activity is he engaged in?) A by someone who can see him] He WRITE letters
- (6) [C=6] He WRITE a letter
- (7) [A: I just talked to my brother on the phone. B: What he do right now? A answers:] He WRITE letters
- (8) [C=7] He WRITE a letter
- (9) [A: I went to see my brother yesterday. B: What he DO? (=What activity was he engaged in?)] He WRITE letters
- (10) [C=10] He WRITE a letter
- (11) [A: I talked to my brother on the phone yesterday. B: What he DO? (=What activity was he engaged in?)] He WRITE letters
- (12) [C=11] He WRITE a letter
- (13) [A: When you visited your brother yesterday, what he DO after you had dinner? ANSWER:] He WRITE letters
- (14) [C=13] He WRITE a letter
- (15) [Q: What your brother DO if you don't go to see him today, do you think? A:] He WRITE a letter (to me)
- (16) [Q: What your brother DO when we arrive, do you think? (=What activity will he be engaged in?)] He WRITE letters
- (17) [C=16] He WRITE a letter
- (18) [Q: What your brother usually DO after breakfast? A:] He WRITE letters
- (19) [C=18] He WRITE a letter
- (20) [Q: What your brother usually DO after breakfast last summer? A:] He WRITE letters
- (21) [C=20] He WRITE a letter
- (22) [Q: What are you planning to do right now? A:] I WRITE letters
- (23) [C=22] I WRITE a letter
- (24) [Neither A nor B can see B's brother. A: What he DO right now, do you think? (=What activity is he engaged in?)] He WRITE letters (I think so because he does that every day at this time)
- (25) [A: My brother works at an office. B: What kind of work he DO?] He WRITE letters
- (26) [A: Last year, my brother worked at an office. B: What kind of work he DO there?] He WRITE letters
- (27) [A: My brother has got a new job. He'll start tomorrow. B: What kind of work he DO there?] He WRITE letters
- (28) [Talking of what happened yesterday] While my brother WRITE the letter, I WAIT in the garden
- (29) [Q: Did your brother finish the letter quickly? A:] (No,) he WRITE the letter slowly
- (30) [Talking of the water in a lake which is visible to the speaker and the hearer:] (The water is usually warm, but today) it BE COLD
- (31) [Of a visible lake, what the water is usually like] It BE COLD
- (32) [Of a visible lake, in which the speaker swam yesterday] (Today the water is warm, but yesterday) it BE COLD
- (33) [Of a visible lake] (The first time I swam in this water many years ago) it BE COLD
- (34) [Of a visible lake, said in the summer] (Usually the water is warm, but this summer) it BE COLD
- (35) [C=34] (Usually the water is warm, but last summer) it BE COLD
- (36) [It's no use trying to swim in the lake tomorrow] The water BE COLD (then)
- (37) [Q: Do you know my brother?] (Yes,) I MEET him (so I know him)
- (38) [C=37] (Yes,) I just (=a couple of minutes ago) MEET him

- (39) [C=37] (Yes,) I MEET him (once) several years ago
- (40) [C=37] (Yes,) I MEET him often (up to now)
- (41) [C=37] (No,) I not MEET him (in my life)
- (42) [Q:] You MEET my brother (at any time in your life until now)?
- (43) [Q: Did you know my father, who died last year?] (Yes,) I MEET him (at least once)
- (44) [C=43] (Yes,) I MEET him (several times, now and then)
- (45) [Assuming that B was going to meet A's brother, A asks:] You MEET my brother (yesterday, as was planned)
- (46) [Answer to (45):] (Yes,) I MEET him (yesterday, as was planned)
- (47) [Answer to (45):] (No,) I not MEET him (yesterday, as was planned)
- (48) [Q: When you came to this place a year ago, did you know my brother?] (Yes,) I MEET him (at least once before I came here)
- (49) [C=48] (Yes,) I MEET him (just before I came here)
- (50) [Q: Did you know my father, who died last year?] (No,) I not MEET him (at any time)
- (51) [Q: When you came to this place a year ago, did you know my brother?] (No,) I not MEET him (before I came here)
- (52) [C=51] (No,) I not MEET him (before I came here but I met him later)
- (53) [A: I want to give your brother a book to read, but I don't know which. Is there any of these books that he READ already? B:] (Yes,) he READ this book
- (54) [A: It seems that your brother never finishes books.] (That is not quite true.) He READ this book (=all of it)
- (55) [Q: Your brother DO what his teacher told him to do today?] (Yes,) he READ (all of) this book (as he was told)
- (56) [Q: Is the king still alive? A:] (No,) he DIE
- (57) [A: Have you heard the news? B: No, what happened? A:] The king BE KILLED (alt: They KILL the king)
- (58) [Q: Do you think the king will go to sleep? A:] (Yes,) he BE TIRED
- (59) [Looking out of the window, seeing that the ground is wet] It RAIN (not long ago)
- (60) [The police are investigating a burglary. Seeing an open window and footprints beneath it, the police inspector says:] The thief ENTER the house by this window
- (61) [It is cold in the room. The window is closed. Q:] You OPEN the window (and closed it again)?

- (62) [Answer to (61):] (Yes,) I OPEN the window
- (63) [Answer to (61):] (No,) I not OPEN the window
- (64) [Child: Can I go now? Mother:] You BRUSH your teeth?
- (65) [A returns home after having been away for a while. B asks:] What you DO?
- (66) [Answer to (65):] I BUY food
- (67) [Q: What did you find out when you came to town yesterday? A:] The king DIE
- (68) [C=67] The king BE KILLED
- (69) [Q: Why is it so cold in the room? The window is open but the person who asks does not know. The person who opened the window answers:] I OPEN the window
- (70) [Q: Has this house always been red? A:] (No, earlier) the house BE WHITE
- (71) [Talking about the speaker's habits: I like to be up early.] I RISE at six in the morning (alt: at dawn)
- (72) [This week I have to go to work early.] I RISE at six in the morning (alt: at dawn)
- (73) [Q: What kind of sound do cats make?] They MEOW
- (74) [Q: What do your cats do when they are hungry?] They MEOW
- (75) If you tease a cat, it MEOW
- (76) [Q: Do cats bark?] (No), they not BARK
- (77) Whatever you TELL him, he not ANSWER
- (78) Whatever you PUT into this bag, it not BREAK
- (79) If you PUT a stone into this bag, it BREAK
- (80) Even if you PUT a stone into this bag, it not BREAK
- (81) [Q: What HAPPEN if I eat this mushroom?] You DIE
- (82) (According to the contract) we not WORK tomorrow
- (83) [Father to child:] (Please do not disturb me), I WRITE a letter
- (84) [Q: Your brother WRITE a letter right now? (=Is that the activity he is engaged in?)] (No,) he not WRITE a letter (he's asleep)
- (85) [Q: What your brother DO right now? (=What activity is he engaged in?) A:] He SIT in a chair (and) READ a book
- (86) [C=85] He EAT bread and DRINK water
- (87) [Q: What your brother DO after breakfast (yesterday)? A:] He GO to the market and/to BUY some apples

- (88) [Q: What the boy's father DO when the boy came home (yesterday)? A:] He BEAT him and KICK him (several times)
- (89) [Q: Did you find your brother at home? A:] (No, we did not.) He LEAVE (before we arrive)
- (90) [C=89] (No, we did not, we were very unlucky.) He LEAVE (just before we came)
- (91) [Q: What your brother's reaction BE when you gave him the medicine (yesterday)?] He COUGH once
- (92) [C=92] He COUGH twice
- (93) [C=92] He COUGH seven times
- (94) [C=92] He COUGH many times
- (95) [C=92] He COUGH for an hour
- (96) [C=92] He COUGH often
- (97) [Q: Why do you think your brother has caught a cold?] He COUGH often
- (98) [Q: Why did you think yesterday that your brother had caught a cold?] He COUGH often
- (99) [Q: How long did it take for your brother to finish the letter?] He WRITE the letter in an hour
- (100) [The boy's father sent him a sum of money some days ago and it arrived yesterday] When the boy GET the money, he BUY a present for the girl
- (101) [Last year, the boy's father sent him a sum of money] When the boy GET the money, he BUY a present for the girl
- (102) [The boy used to receive a sum of money now and then] When the boy GET the money, he BUY a present for the girl
- (103) [The boy is expecting a sum of money] When the boy GET the money, he BUY a present for the girl
- (104) [The boy thinks that he will perhaps get a sum of money] If the boy GET the money, he BUY a present for the girl
- (105) [The speaker knows the boy was expecting money, but he doesn't know if he got it] If the boy GET the money (yesterday), he BUY a present for the girl
- (106) [The speaker knows the boy was expecting money and that he did not get it] If the boy GET the money (yesterday), he BUY a present for the girl
- (107) [Talking to someone who is leaving in a while] When you RETURN, I WRITE this letter (=I FINISH it already at that time)

- (108) [Said as an order by a teacher leaving the classroom] When I RETURN, you WRITE this assignment (=You FINISH it by then)
- (109) [Assuming that the speaker's brother is trustworthy and speaking of the water in a lake which is not visible to the speaker and the hearer] My brother SAY (right now) that the water BE COLD
- (110) [Of the water in a lake which is not visible to the speaker and the hearer] My brother SAY (right now) that the water BE COLD (but I don't believe him)
- (111) [C=110] My brother SAY (right now) that the water BE COLD (yesterday, but I don't believe him)
- (112) [C=110] My brother SAY (yesterday) that the water BE COLD (yesterday, but I think he was wrong)
- (113) [C=110] My brother SAY (yesterday) that the water BE COLD (the day before yesterday, but I think he was wrong)
- (114) [C=110] My brother SAY (yesterday) that the water BE COLD (today, but he turned out to be wrong)
- (115) [C=110] My brother THINK (right now) that the water BE COLD (today, but he is wrong)
- (116) [C=110] My brother THINK (yesterday) that the water BE COLD (yesterday, but he was wrong)
- (117) [C=110] My brother KNOW (now) that the water BE COLD (today)
- (118) [C=110] My brother KNOW (yesterday) that the water BE COLD (today)
- (119) [C=110] My brother BELIEVE (yesterday) that the water BE COLD (usually)
- (120) [C=110] He FEEL (right now) that the water BE COLD
- (121) [C=110] (Yesterday when my brother went into the water) he FEEL that the water BE COLD
- (122) [Of the water in an invisible lake - the speaker knows that the water is in fact cold] My brother not THINK (right now) that the water BE COLD (=he thinks it is warm)
- (123) [C=110] My brother DOUBT (right now) that the water BE COLD (=he suspects it is warm)
- (124) [C=110] My brother HOPE (right now) that the water BE COLD
- (125) [Uttered as a promise] I PROMISE to COME to you tomorrow
- (126) [Uttered in a naming ceremony (for NAME, substitute any suitable verb (e.g. CHRISTEN), for X, substitute any suitable proper name (e.g. JOHN)] I NAME this child X
- (127) [Looking at a broken cup, angrily:] Who BREAK this cup?

- (128) [Looking at a house] Who BUILD this house?
- (129) [Looking at a picture of a house which has been torn down] Who BUILD this house?
- (130) [Looking at a house, recently painted] Who PAINT this house?
- (131) You MUST GO to bed before you GET tired (today)
- (132) (Yesterday evening) I GO to bed before my brother COME home
- (133) [The speaker has just seen the king arrive (no one had expected this event)] (Have you heard the news?) The king ARRIVE
- (134) [A person who has heard (133) but not seen the event says:] (Have you heard the news?) The king ARRIVE
- (135) [The king has been expected for weeks. The speaker has just seen him:] The king ARRIVE
- (136) [A person who has heard (135) but not seen the event says:] The king ARRIVE
- (137) When I COME home (yesterday), he WRITE two letters (=first I came and then he wrote the letters)
- (138) When I COME home (yesterday), he WRITE two letters (=he finished writing them just before I came)
- (139) When I COME home (yesterday), he WRITE two letters (=that is what he accomplished during my absence)
- (140) When I COME home (yesterday), he WRITE two letters (=that is the activity he was engaged in)
- (141) [Conversation takes place in the afternoon. The market referred to is assumed to be situated at a considerable distance from the place where the conversation takes place Q: Do you know my brother?] (Yes,) I MEET him at the market this morning
- (142) [C=141] (Yes,) I MEET him at the market yesterday
- (143) [Conversation takes place in the afternoon: Do you know my brother?] (Yes,) I MEET him here this morning
- (144) [C=143] Do you know my brother?] (Yes,) I MEET him here yesterday
- (145) [Traveller to local:] If you SHOW me the way, I GIVE you money
- (146) [Mother to child:] If you not STOP PLAY with that ball, I TAKE it away
- (147) [Standing in front of a house: Who BUILD this house?] My brother BUILD this house
- (148) [(Of a coughing child:) For how long has your son been coughing?] He COUGH for an hour

- (149) [A knows that B was going to meet A's brother but not when. A:] You MEET my brother (yet)?
- (150) [Answer to (149):] (Yes,) I MEET him.
- (151) [Answer to (149):] (No,) I not MEET him
- (152) [Said by a young man] When I GROW old, I BUY a big house
- (153) [The speaker has just seen the king arrive (earlier than was expected):] The king ARRIVE already
- (154) [The king is expected to arrive] The king not ARRIVE yet
- (155) [Q: Has your brother finished the letter?] (No,) he still WRITE it
- (156) [Q: What did your brother say yesterday when you asked him if was busy?] He SAY that he WRITE letters

Part B - connected texts

- (B1) [Do you know what happened to me yesterday?] (161) I WALK in the forest. (162) Suddenly I STEP on a snake. (163) It BITE me in the leg. (164) I TAKE a stone and THROW at the snake. (165) It DIE.
- (B2) [I'll tell you what happened to me once when I was a child.] (166) I WALK in the forest. (167) Suddenly I STEP on a snake. (168) It BITE me in the leg. (169) I TAKE a stone and THROW at the snake. (170) It DIE.
- (B3) [Do you know what happened to my brother yesterday? I saw it myself.] (171) We WALK in the forest. (172) Suddenly he STEP on a snake. (173) It BITE him in the leg. (174) He TAKE a stone and THROW at the snake. (175) It DIE.
- (B4) [Do you know what happened to my brother yesterday? He told it himself.] (176) He WALK in the forest. (177) Suddenly he STEP on a snake. (178) It BITE him in the leg. (179) He TAKE a stone and THROW at the snake. (180) It DIE.
- (B5) [Once upon a time there was a man. This is what happened to him one day.] (181) He WALK in the forest. (182) Suddenly he STEP on a snake. (183) It BITE him in the leg. (184) He TAKE a stone and THROW at the snake. (185) It DIE.
- (B9) [The speaker is right back from a walk in the forest: Do you know what just happened to me?] (186) I WALK in the forest. (187) Suddenly I STEP on a snake. (188) It BITE me in the leg. (189) I TAKE a stone and THROW at the snake. (190) It DIE.
- (B6) [I'll tell you what happens to me sometimes when I am walking in the forest.] (191) I SEE a snake. (192) I TAKE a stone and THROW at the snake.

- (B7) [I'll tell you what happened to me sometimes, when I was a child and was walking in the forest.] (193) I SEE a snake. (194) I TAKE a stone and THROW at the snake.
- (B8) [I'll tell you what I see in the window right now.] (195) A boy and a girl PLAY in the street. (196) (Right now) the boy TAKE a ball and THROW it to the girl. (197) The girl THROW it back.

Index

- accidence, 12
 accidence category, 11-15, 188
 adjectives, semantics of, 17
 adverbial phrases, specifying
 length of time, 76
 Afrikaans (AF), 104, 112, 166;
 gaan+infinitive construction in,
 112; past time reference in, 28
 Afro-Asiatic (language-group),
 154
 Aghem, remoteness distinctions
 in, 126
 Akan (AK), 70, 71, 90, 96, 100,
 104, 129, 130, 179;
 FRAMEPAST in, 147; marking
 of two-way actions, 149
 Aksu, A., 151
 Aktionsart, 26, 27
 Alawa (AW), 70, 71, 101, 104, 115
 algebra, Boolean, 7, 15, 17
 Allwood, J., 13
already, 129
 Altaic (language-group), 157
 Amharic (AM), 71, 115, 130, 145,
 156
 Andean-Equatorial (language-
 group), 158
 Andean-Quechumaran (language-
 group), 158
 Andersson, S. G., 27
 Angenot, J.-P., 125
 Appleby, L. L., 136
 Arabic, 155-6
 Arabic, Modern Standard (AB),
 70, 71, 105, 110, 115, 155;
 Perfective in, 80; pragmatic
 particles in, 53; TMA paradigm,
 83;
 Arabic, Tunisian (AR), 70, 71,
 108, 115, 155
 Arnauld, A., 125
 areal-genetic groups, 43
 aspect, notion of, 23, 24; vs
 Aktionsart, 26
 aspectual potential of verbs, 26
 Australian (language-group), 159
 Austronesian (language-group),
 160
 Azerbaijani (AZ), 70, 71, 101,
 102, 105, 110, 111, 115, 130, 131,
 144, 159-60, 157; Aorist in, 111;
 development of PROG into
 Simple Present 93; TMA
 paradigm, 157
 Bandjalang (BA), 70, 71, 90, 101,
 104, 159
 Bantu (language-group), 175, 185
 Beja (BJ), 70, 71, 104, 110, 120,
 130, 149, 154-5; past time
 reference in, 28, 120
 Bell, A., 37, 43
 Bengali (BE), 90, 101, 102, 104,
 115, 125, 130, 145
 Berlin, B., 33
 Bickerton, D., 119, 167
 Bierwisch, M., 191
 boundedness, 29; as determinant
 of aspect, 76; vs boundness 192
 Bugis Makassar (BG), 104, 110,
 130, 149, 162
 Bulgarian (BU), 70, 71, 86, 104,
 130, 167, 172; Aorist/Non-
 Aorist, 85; development of
 QUOT in, 152; habitual contexts
 in, 78; Imperfective Aorist in,
 77; Imperfect-Aorist distinction
 in 28; Perfect in, 137; Perfective/
 Imperfective, 85; QUOT in, 151
 C: see contingency coefficient
 Carlson, G., 98

- Catalan (CA), 70, 71, 91, 104, 110, 115, 130, 145, 170, 172; *acabar* de construction in, 127; hodiernal:non-hodiernal distinction in, 125
- category, 20, 21; Boolean, 63; complex, 67; default, 19; derivational, 185; focused, 4; grammatical, 12, 16, 20, 21, 22; vs lexical, 26; inflectional, 12, 186; language-specific, 34; morphological, 12, 184; notional 21; periphrastical, 184; unmarked, 19
- category-types, cross-linguistic, 34
- Caucasian (language-group), 124, 164
- Cebuano (CE), 70, 71, 98, 139, 160
- centrality of TMA categories; criteria for, 188
- centre of TMA system, 184
- Chia, N., 124
- Chinese (Mandarin) (CH), 70, 71, 90, 105, 139, 180; *guo* construction in, 141, 143
- Chomsky, N., 31
- coding of categories, 46; superficiality principle of, 46, 53
- Colarusso, J., 124
- Cole, D. T., 114
- combinatorial explosion, 47
- comparisons, cross-linguistic, 54
- competence, innate, 31
- computer analysis, 50
- Comrie, B., 19, 22, 23, 24, 25, 27, 30, 34, 74, 75, 81, 82, 94, 111, 132, 133, 142
- conceptual space, dimensions of, 34; innateness of, 34
- CONCL: *see* CONCLUSIVE
- CONCLUSIVE (CONCL), 95
- conditions of use, 14
- construction type, dynamic, 28; stative, 28
- contingency coefficient, 54
- core of TMA system, 23
- Creole languages, 119, 167
- cross-linguistic vs universal, 32
- Cushitic (language-group), 154
- Czech (CZ), 70, 71, 86, 96, 100, 105, 110, 115, 117, 130, 172; habitual contexts in, 78
- Danish, PFCT in, 138
- data base system, 51
- data, choice of, 36; primary, 36; secondary, 36
- Davies, J., 125
- default category, 19
- definitions in science, 20
- Derbyshire, D. C., 121
- derivation vs inflection, 185
- distance, R-E, 121; S-E, 121; temporal, 120; objective and subjective measures, 121
- distribution, ideal, 55
- dominance, 8-9, 23
- Dravidian (language-group), 164
- durative, 72
- DYNAMIC, 29
- English (EN), 31, 90, 101, 102, 104, 111, 115, 130, 144, 166; extended uses of PROG in, 93; Perfect Progressive in, 152; Progressive in, 28
- errors, sources of, 48
- Eskimo, West Greenlandic (GL), 98, 100, 104, 149, 165; quotatives in, 192; suffix *-riataar-* in, 95
- Eskimo-Aleut (language-group), 164, 165
- Estonian (ES), 71, 108, 115, 130, 145, 181; future time reference in, 109; genitive in, 69, 89; modus obliquus in, 151; Partitive in, 143
- events, generic, 27; individual, 27
- Ewondo, 125
- EXPER: *see* EXPERIENTIAL
- EXPERIENTIAL (EXPER), 129, 139-44; categories, 139; distribution, 140; vs PFCT, 140; prototypical occurrences of, 140; repeatability condition on, 143; semantics, 140
- extensional vs intensional 9
- Fant, G., 31
- features, binary, 15, 17
- Finnish (FI), 71, 91, 108, 115, 130, 145, 181; Accusative in, 69, 89; Partitive in, 143
- Fitzroy Crossing Kriol (FC), 90, 104, 115, 166; *-bat* form in, 94
- focus, 4, 10; secondary, 11
- focusing, 3
- frame, temporal, 30; past 146-7

- FRAMEPAST, 144, 148-9
- French (FR), 70, 71, 90, 104, 111, 115, 145, 170; *aller+inf* construction in, 112; disappearance of remoteness distinctions in, 125
- frequency, as a criterion of centrality, 188
- Friedrich, P., 72, 80
- FUTi, 110
- FUTURE (FUT), 103-111, 190; categories, 104; distribution, 106; languages without, 108; obligatoriness of, 109; semantics, 104
- future time reference, 23, 103
- fuzziness, 4
- fuzzy set theory, 4, 9
- Galla: *see* Oromo
- gender, 191; natural, 12
- generalizations, cross-linguistic, 31
- generic sentences, 98
- generics, 95
- Georgian (GO), 70, 71, 96, 100, 104, 115, 130, 131, 164; aspect in, 87
- German (GE), 90, 96, 100, 104, 110, 115, 130, 145, 166; dialects, 167
- Germanic (language-group), 166
- Gesamtbedeutung, 74
- Greek, Classical, 169; augment in, 83
- Greek, Homeric, Aorist in, 80
- Greek Modern (GR), 70, 71, 104, 115, 130, 144, 169; PFV in, 80
- Grice, H. P., 11
- Gricean principles, 11, 14
- Groot, C. de, 87
- Guarani (GA), 91, 96, 100, 104, 110, 120, 159; future time reference in, 192
- Gur, 177
- HAB: *see* HABITUAL
- HABG: *see* HABITUAL-GENERIC
- HABITUAL (HAB), 96-7, 190; categories, 97; distribution, 97; marking of time reference in, 99; prototypical occurrences of, 97
- habitual 95, vs generic, 96; vs iterative, 96
- HABITUAL-GENERIC (HABG), 95, 97-100, 194; categories, 98; distribution, 98; marking of time reference in, 99; prototypical occurrences of, 99
- HABITUAL-PAST (HABPAST), 95, 100-2, 189; categories, 101; distribution, 101; prototypical occurrences of, 101; semantics of, 101
- HABPAST: *see* HABITUAL-PAST
- Halle, M., 31
- Haltorf, B., 18
- Hawaiian (HA), 70, 71, 94, 162-3
- Hebrew, Biblical, wa-construction in, 114
- Hebrew, Modern (HE), 104, 115, 155
- HEST: *see* HESTERNAL
- HESTERNAL (HEST), 126
- Hindi: *see* Hindi/Urdu
- Hindi/Urdu (HI), 71, 90, 104, 130, 145, 168, 169; development of PROG into IPFV, 93
- hit rate, 55, 59
- Hixkaryana, remoteness system of, 121
- HODIERNAL (HOD), 125
- HOD-PAST, 135
- Hoffmann, C., 88
- homomorphism, 7
- Hopper, P., 89, 161
- horizon of interest, 124
- Hungarian (HU), 71, 100, 104, 115, 181; aspect in, 86
- Hyman, L., 126
- Hymes, D., 121, 125, 126
- IMMEDIATE PAST (IMMED-PAST), 127
- imperative, 26
- IMPERFECT (IMPFT), 117
- IMPERFECTIVE (IPFV), 69, 190; general factual use of, 75; vs PROG, 92
- implicatures, conversational, 11, 25; conventionalization of, 11
- impreciseness, 30; multi-dimensional vs one-dimensional, 6; in science, 2
- Indo-European (language-group), 42, 166; classic TMA system of, 81

- Indo-Iranian (language-group), 167
 Indonesian (IN), 90, 105, 130, 139, 160; PFCT and EXPER in, 142
 inferentiality, 152
 inflection vs derivation, 185
 informants, choice of, 44; role of, 43
 Inoue, K., 141, 142
 intensional vs extensional, 9
 intention, 105
 investigator, role of, 43
 Iroquois (language-group), 173
 Isekiri (IS), 90, 98, 100, 104, 130, 139; lexical taboos in, 193
 Italian (IT), 70, 71, 90, 104, 110, 144, 170, 179
- Jakobson, R., 15, 31
 Japanese (JA), 71, 108, 110, 139, 149, 158; PFV in, 80; pragmatic particles in, 53; *-ta koto ga aru* construction in, 141, 142; *-te shimau* construction in, 95
 Javanaud, P., 124
 Javanese (JV), 90, 104, 130, 139, 161; PFCT and EXPER in, 142
 Jespersen, O., 21, 138
 Johnson, B., 114
 Johnson, M., 136
- Kamba, remoteness system of, 121; HOD-PAST vs PFCT in, 136
 Kam-Tai (language-group), 174
 Kammu (KM), 96, 104, 130, 174; *hòoc* construction in, 57, 59
 Kamp, J. A. W., 17
 Karaboro (KB), 71, 94, 104, 113, 114, 117, 177; dependency marker in, 114; particle *yaa* in, 129
 Kartvelian (language-group), 164
 Katenina, T. E., 147
 Katz, J. J., 191
 Kay, P., 33
 Keenan, E., 35
 Kiksht, remoteness system of, 121, 125, 126
 Kikuyu (KI), 71, 94, 108, 113, 125, 129, 136, 175
 kinship distinctions, marking of, 124
 Kobon, hodiernal:non-hodiernal distinction in, 125
- Kom, remoteness system of, 121
 Kopchevskaya, M., 120
 Kós-Dienes, D., 154, 193
 Kriol: *see* Fitzroy Crossing Kriol
 Kučera, H., 102
 Kurdish (KU), 70, 71, 104, 115, 144, 149, 168; development of QUOT in, 152
 Kwa, 179
- Lancelot, C., 125
 language acquisition, 191
 language change, 191
 Latin (LN), 70, 71, 104, 115, 144, 170; imperfectum de conatu in, 89
 lattice, 7
 Latvian, aspect in, 86
 Li, C. N., 135, 181
 Limouzi (LI), 70, 71, 90, 104, 115, 130, 145, 170, 171; remoteness distinctions in, 124
 Lithuanian, aspect in, 86
 Ljung, M., 94
 Lonkundo, 136
 Lounsbury, F., 173
 Lundberg, L. J., 64
- Maltese (ML), 70, 71, 90, 104, 110, 115, 156
 Maori (MR), 90, 98, 108, 125, 130
 Maran (language-group), 160
 Margi, 88
 markedness, 19; vs centrality, 189; paradoxical, 19
 marking type, 34
 Martin, S., 95
 maxim of quantity, 14
 maxim of relation, 14
 McCawley, J. D., 132, 142, 147
 McCoard, R., 133, 136
 meaning, basic, 9; inherent aspectual, 26, 27; primary, 9; secondary, 10
 Mon-Khmer, 174
 Mønnesland, S., 78
 mood, 23; categories of, 53; notion of, 23, 25, 26; and pragmatics, 5
 Morolong, M., 124
 Muli, M. G., 121
- NARR categories, 113
 narrative context, 112; discourse, 112

- narrativity, 112; and remoteness, 126
 native speaker requirement, 44
 Nedjalkov, 133
 negated clauses, 47
 Niger-Congo (language-group), 42, 175
 nouns, mass-count distinction, 76
 'now-tense', 95, 176
 Nyuni (language-group), 176
- obligatoriness of expression, 189
 Omamor, A. P., 193
 Oneida (ON), 71, 101, 104, 105, 173; FRAMEPAST in, 148; marking of two-way actions, 149
 operators vs features, 16-18
 opposition, equipollent, 16, 72; privative, 16
 Oromo (Galla) (OR), 71, 94, 130, 145, 163, 154, 155
 ostension, 20
- Pama-Nyungan, 160
 parameters in universal grammar, 31
 passives, 47
 PAST, 63, 116, 189; categories, 115; marking-types, 184
 PAST-PROG, 94
 PASTd, 120
 PASTi, 118; distribution, 118; prototypical occurrences of, 118
 PASTn, 119
 PASTs, 119
 past, hesternal, 126; hodiernal, 126, 136
 Past Perfect: *see* PLUPERFECT
 perdurative, 77
 PERFECT (PFCT), 56, 60, 62, 63, 129-139, 189, 190; categories, 130; and definiteness, 137; and recency, 135; vs EXPER, 140; 'extended now' theory of, 136; inferential uses of, 152, 153; interpretations of, 131, 132; marking of, 129; and narrativity, 137; vs PFV, 62, 138; vs QUOT, 152
 perfect, 'hot news', 132; existential, 132; experiential, 132, 142; of persistent situation, 132, 136; of recent past, 132; of result, 132, 133; universal, 136
- PERFECT-QUOTATIVE (PFCTq), 130, 152
 PERFECTIVE (PFV), 23, 62, 69-89, 189; prototypical occurrences of, 77; restrictions on time-reference, 79
 perfectivity, 23, 69; relation to tense, 81; 'totality' view of, 74
 PERFECTIVE:IMPERFECTIVE (PFV:IPFV) opposition, 69-89; in habitual contexts, 78; semantics of, 73
 performatives, 81
 periphery, 4; of TMA system, 23, 184;
 Persian (PE), 70, 71, 90, 104, 110, 115, 130, 131, 145, 168
 Pettersson, Th., 18
 PFCT: *see* PERFECT
 PFCTq: *see* PERFECT-QUOTATIVE
 PFV: *see* PERFECTIVE
 PFV:IPFV: *see* PERFECTIVE:IMPERFECTIVE
 PFVd, 70, 84
 PLPFCT: *see* PLUPERFECT
 PLUPERFECT (PLPFCT), 144-7, 189, 190; as remote past, 147; categories, 144; in counterfactual sentences, 146; distribution, 145; marking past temporal frames, 147; prototypical occurrences of, 145
 point of event (E), 29
 point of reference (R), 29
 point of speech (S), 29
 Polish (PO), 70, 71, 86, 105, 110, 115, 117, 172; habitual contexts in, 78
 Polynesian (language-group), 162, 163
 Portuguese (PG), 70, 71, 90, 104, 105, 115, 130, 144, 171, 172
 post-hodiernal future, 126
 pragmatics, vs semantics, 3
 PRED: *see* PREDICTIVE
 PREDICTIVE (PRED), 110, 190; categories, 110; distribution, 111; prototypical occurrences of, 111
 pre-hodiernal, 125
 predicates, nominal, 28; verbal, 28
 prediction, 104
 prefixation, 187; vs suffixation, 185

- present, historical, 113; reportive, 72, 81
PRESENT-PROGRESSIVE (PRES-PROG), 94
PROG-PAST, 94
PROGRESSIVE (PROG), 25, 56, 73, 90-5, 189; categories, 90; distribution, 91; extended uses of, 93; vs **IPFV**, 92; marking-types, 185; problematic cases of, 94; prototypical occurrences of, 92; semantics, 91; time-restricted cases of, 94
PROSP: see PROSPECTIVE
PROSPECTIVE, 111, 190
 prototype, 4
Punjabi (PU), 70, 71, 91, 104, 130, 145, 168; development of **PROG** into **IPFV**, 93
Putonghua: see Chinese, 180

Quechua (QU), 104, 125, 149, 158, 159; 'Progressive' in, 94; **PASTn** in, 120
Quine, W. V. O., 1, 34
QUOT: see QUOTATIVE
QUOTATIVE (QUOT), 129, 149, 190; categories, 149; in fiction, 151; in indirect speech, 151; in inferential contexts, 152; prototypical occurrences of, 149; vs **PFCT**, 152; semantics, 149;

Rafferty, E., 161
 ranking list of category
 distribution, 57; bias of, 66
 ranking list method, limitations of, 62
Reichenbach, H., 29-31, 112; theory of tense, 29
 relative precedence of categories, 188
 relativism, linguistic, 32
REM-PAST: see REMOTE-PAST
REMOTE-PAST (REM-PAST), 175 remoteness, 120; categories, 127; vs narrativity, 126
 remoteness categories, 127
 result, notion of, 134
RESUL: see RESULTATIVE
RESULTATIVE (RESUL), 133-5, 191; and passive voice, 135
Romance, 170
Romanian (RO), 70, 71, 104, 110, 115, 144, 171, 172
Russian (RU), 70, 71, 86, 105, 110, 115, 172; habitual contexts in, 78; imperfectivity-perfectivity distinction, 27; marking of past temporal frames in, 148; marking of two-way actions, 149; overview of TMA system, 84; **PFV:IPFV** in, 75; 'pluperfect' in, 191; restrictions on **PFV** in, 80; TMA paradigm, 85

 sampling, accessibility biases in, 38; areal bias in, 37; genetic bias in, 37
Sapir, E., 31
 self-organizing systems, 191
 semantics, vs pragmatics 3
Semitic (language-group), 42, 155; typical TMA system, 82
Seneca (SE), 70, 71, 98, 101, 102, 104, 173
Serbo-Croatian, habitual contexts in, 78
Setswana, NARR in, 114
 sharpening, 5
Siamese: see Thai
Siegel, S., 54
Sino-Tibetan (language-group), 180
 situation, dynamic, 28
 situations, taxonomy of, 27
Slavic (language-group), 172; aspect in, 69, 84; derivational character of aspect, 85; **IPFV** in, 143
Slobin, D., 151
Slovak, habitual contexts in, 78
Slovene, habitual contexts in, 78
Sorbian, habitual contexts in, 78
Sotho (SO), 71, 90, 104, 113, 115, 139, 175; auxiliary se in, 176; **NARR** in, 114; **PFV** in, 80; remoteness system of, 124
South Celebes (language-group), 162
Spanish (SP), 70, 71, 90, 104, 105, 110, 115, 130, 145, 159, 171; *acabar de* construction in, 127; hodiernal: non-hodiernal distinction in, 125
 species, reference to, 10
 state, 28
STAT: see STATIVE
STATIVE (STAT), 29

- 'still-tenses', 176
Stojanov, S., 151
 subject, notion of, 35
 subjunctive, 26
 suffixation, vs prefixation, 185
Sundanese (SD), 90, 108, 130, 139; bade construction in, 112; **PFCT** and **EXPER** in, 142
Swedish (SW), 64, 90, 96, 100, 104, 110, 115, 130, 145, 161 166; future constructions, 107, 108; inferential **PFCT** in, 152; Perfect in, 137; resultative construction in, 134

Tahitian, 162, 163
Talmoudi, F., 53
Tamil (TA), 90, 115, 130, 145, 165-6; 'future tense' in, 104; *vida* construction in, 95
 tense, absolute, 25; categories, 103-127; notion of, 23-5; relative, 25 tense logic, 18
Thai (TH), 90, 104, 130, 139, 156, 174
Thelin, N. B., 18, 192
Thompson, S., 89
Tigrinya (TG), 94, 104, 110, 115
 time, situation-external vs situation-internal, 25
TMA category, 1, 22; 'major' 52
TMA Questionnaire, 37, 44; cultural bias in, 48; problems in constructing, 47
TMA system, 22; basic units of, 33
 transitivity and aspect, 89
 translation method, unreliability of, 49
 truth-conditions, 13
Tunisian: see Arabic
Tupi (language-group), 159
Turkic (language-group), 157

Turkish (TU), 70, 71, 104, 111, 115, 149, 157; Aorist in, 111; development of **PROG** into Simple Present, 93; development of **QUOT** in, 152; **QUOT** in, 151
 two-way action, 149
 typology morphological, 187

Ultan, R., 121, 193
 universals, absolute, 31; implicational, 31; in linguistics, 31; statistical, 31
Uralic (language-group), 180
Urdu: see Hindi/Urdu

 vagueness, 4
 variation, inter-vs intralinguistic, 64
Vendler, Z., 191
 verbs, dynamic vs stative, 28; **Vinogradov, V. V.**, 76
Volta-Comoe (language-group), 179

Weinreich, U., 191
Welmers, W., 136 West
 Indonesian (language-group), 160
Whiteley, W. H., 121
Wichser, M., 193
Wolff, J. U., 161
Wolof (WO), 71, 96, 98, 100, 104, 120, 130, 180; **PASTn** in, 120

 'yet-tenses', 176
Yiddish, 167
Yoruba (YO), 90, 96, 104, 105, 110, 130, 179

Zulu (ZU), 71, 113, 125, 175; **PFV** in, 80