CAMBRIDGE TEXTBOOKS IN LINGUISTICS

Typology and Universals

Second Edition

William Croft

Typology and Universals

Comparison of the grammars of human languages reveals systematic patterns of variation. Typology and universals research uncovers those patterns to formulate universal constraints on language and seek their exploration. In this essential textbook, William Croft presents a comprehensive introduction to the method and theory used in studying typology and universals. The theoretical issues discussed range from the most fundamental – on what basis can the grammars of diverse languages be compared? – to the most abstract – what is the role of functional and historical explanations of language universals? The book provides students and researchers with extensive examples of language universals in phonology, morphology, syntax and semantics.

This second edition has been thoroughly rewritten and updated to reflect advances in typology and universals in the past decade, including: new methodologies such as the semantic map model and questions of syntactic argumentation; discussion of current debates over deeper explanations for specific classes of universals; and comparison of the typological and generative approaches to language.

WILLIAM CROFT is Professor of Linguistics at the University of Manchester. His books include *Studies in typology and diachrony for Joseph H. Greenberg* (edited with Keith Denning and Suzanne Kemmer, 1990), *Typology and universals* (1990), *Syntactic categories and grammatical relations: the cognitive organization of information* (1991), *Explaining language change: an evolutionary approach* (2000) and *Radical Construction Grammar: syntactic theory in typological perspective* (2001).

CAMBRIDGE TEXTBOOKS IN LINGUISTICS

General Editors: S. ANDERSON, J. BRESNAN, B. COMRIE, W. DRESSLER, C. J. EWEN, R. HUDDLESTON, R. LASS, D. LIGHTFOOT, P. H. MATTHEWS, S. ROMAINE, N. V. SMITH, N. VINCENT

In this series

P. H. MATTHEWS *Morphology* Second edition
B. COMRIE *Aspect*
R. M. KEMPSON *Semantic Theory*
T. BYNON *Historical Linguistics*
J. ALLWOOD, L.-G. ANDERSON and Ö. DAHL *Logic in Linguistics*
D. B. FRY *The Physics of Speech*
R. A. HUDSON *Sociolinguistics* Second edition
A. J. ELLIOT *Child Language*
P. H. MATTHEWS *Syntax*
A. RADFORD *Transformational Syntax*
L. BAUER *English Word-Formation*
S. C. LEVINSON *Pragmatics*
G. BROWN and G. YULE *Discourse Analysis*
R. HUDDLESTON *Introduction to the Grammar of English*
R. LASS *Phonology*
B. COMRIE *Tense*
W. KLEIN *Second Language Acquisition*
A. J. WOODS, P. FLETCHER and A. HUGHES *Statistics in Language Studies*
D. A. CRUSE *Lexical Semantics*
F. R. PALMER *Mood and Modality*
A. RADFORD *Transformational Grammar*
M. GARMAN *Psycholinguistics*
W. CROFT *Typology and Universals*
G. G. CORBETT *Gender*
H. J. GIEGERICH *English Phonology*
R. CANN *Formal Semantics*
P. J. HOPPER and E. C. TRAUGOTT *Grammaticalization*
J. LAVER *Principles of Phonetics*
F. R. PALMER *Grammatical Roles and Relations*
B. BLAKE *Case*
M. A. JONES *Foundations of French Syntax*
A. RADFORD *Syntactic Theory and the Structure of English: a Minimalist Approach*
R. D. VAN VALIN, JR. AND R. J. LAPOLLA *Syntax: Structure, Meaning and Function*
A. DURANTI *Linguistic Anthropology*
A. CRUTTENDEN *Intonation* Second edition
J. K. CHAMBERS and P. TRUDGILL *Dialectology* Second edition
C. LYONS *Definiteness*
R. KAGER *Optimality Theory*
J. HOLM *An Introduction to Pidgins and Creoles*
C. J. EWEN and H. VAN DER HULST *The Phonological Structure of Words: an Introduction*
E. GUSSMAN *Phonology*
M. YIP *Tone*
W. CROFT *Typology and Universals* Second edition

Typology and Universals
Second Edition

WILLIAM CROFT

Department of Linguistics
University of Manchester

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521004992

First published 1990
Reprinted 1993, 1996
Second edition 2003
Third printing 2006

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data

Croft, William.
Typology and universals / William Croft. – 2nd edn.
p. cm. – (Cambridge textbooks in linguistics)
Includes bibliographical references and index.
ISBN 0-521-80884-7 – ISBN 0-521-00499-3 (pbk.)
1. Typology (Linguistics) 2. Universals (Linguistics) I. Title. II. Series.
P204 .C7 2002
410′.1 – dc21 2002019247

ISBN 978-0-521-80884-2 hardback
ISBN 978-0-521-00499-2 paperback

Transferred to digital printing 2009

To the Memory of Joseph H. Greenberg (1915–2001)

Contents

Figures

Tables

Preface to the second edition

This second edition of *Typology and universals* is almost completely rewritten from the first edition. Although the number of chapters and much of their content remains the same, many major changes have been made, largely due to the maturing of typology as an approach to language. The most important innovation is the systematic employment of the semantic map model, now widely used in typological research. Also, I have reorganized the material so that typological generalizations and their explanations are now more closely integrated.

Typology has also developed an independent institutional identity in the past decade. There is now a journal, *Linguistic Typology*, and an international association (the Association for Linguistic Typology [ALT]) with biennial conferences. A Max Planck Institute for Evolutionary Anthropology has been established in Leipzig, Germany, with a typologically oriented linguistics section under the direction of Bernard Comrie. These institutional developments also reflect a shift in the center of gravity in typology from the United States to Europe. Major typological studies have been published in the last decade or so by scholars based in Europe (including Russia, long a center of typological research). Some of this shift is reflected in this edition and in the references in the Bibliography.

This edition has benefited from the input of students in ten years of classes in typology at the University of Michigan and the University of Manchester, as well as shorter courses at the LSA Summer Institute in Albuquerque, New Mexico, USA in 1995, the Summerschool of the German Linguistics Society, Mainz, Germany in 1998, and the LOT Winterschool in Leiden, the Netherlands in 2000. I wish to thank all of those students for their input. I have also devised a number of problem sets for use in teaching typology and universals, which have benefited from my students' experiences and difficulties in solving them. These problem sets are not included in this textbook for reasons of space; they can be found at http://lings.ln.man.ac.uk/html/WAC/.

This edition has also benefited from many comments on and reviews of the first edition, and from presentations based on topics now in this edition. I am especially grateful to Bernard Comrie, Sonia Cristofaro, Matthew Dryer and Martin Haspelmath, who read the entire final draft and gave me extensive comments, which

greatly improved the manuscript. None of them bear any responsibility for errors that remain.

Two individuals who played a significant role in my education in typology have died since the first edition was published in 1990. Keith Denning gave me valuable advice in the development of the original textbook, and gave me much advice and support until his untimely death. I owe a deep intellectual and personal debt to him (see Croft 2000). Finally, I would like to reaffirm my deep appreciation and respect for the founder of modern typology, and my teacher, Joe Greenberg. The field is still deeply indebted to his pioneering theoretical work (e.g. 1954; 1957; 1966a; 1966b; 1966c; 1969/1990; 1978b), as can be seen in this textbook. Moreover, he produced some of the major empirical studies in typology, in both morphosyntax (e.g. 1966a; 1966b; 1978c) and phonology (e.g. 1970; 1978a), which still stand as major discoveries of universals of language. All of my own work has been profoundly influenced by him. I dedicated the first edition of this book to him. Sadly, Joe Greenberg died before being able to see the second edition. I dedicate this edition to his memory.

Preface to the first edition

This volume is an introduction to the concepts and methodology of linguistic typology. It complements other introductory volumes on typology, particularly Comrie 1989 and Mallinson and Blake 1981, in that the material is organized by theoretical concept (implicational universal, markedness, prototype) rather than by topic area (word order, grammatical relations, relative clauses, animacy). Also, the range of concepts covered is somewhat broader, mostly because of the need to describe developments in functional–typological explanation and diachronic typology in the last decade. Needless to say, there is some overlap with the aforementioned volumes. From a pedagogical point of view, however, this volume is intended to complement, not supplement, the more topic-oriented introductions. In particular, breadth in theoretical coverage has meant that detailed examples of typological generalizations, complete with qualifications, possible counterexamples and explanations for those counterexamples, could not always be included (though I have tried not to oversimplify examples without at least citing more detailed studies). The material in this volume has been used in courses in conjunction with Comrie 1989, Greenberg 1966a (the original article on word order), Greenberg 1966b (the monograph on markedness) and other articles on more specific topic areas.

I believe that an essential part of any linguistics class, and above all any class on typology, is for the student to encounter one or more 'exotic' languages. For practical reasons, in an introductory typology class this encounter must be somewhat limited. In my course, each student is required to 'adopt' a grammar of an exotic language, from a list of grammars that I considered particularly detailed, careful and thorough (though not always easy to use!). Each student then writes short papers describing a particular aspect of the grammar, such as negative constructions or word order in the noun phrase. Some of these assignments include group efforts in order to give the students a chance to compare languages on their own.

These essentially descriptive assignments are not as easy as they may seem, as anyone who has actually done typological research using grammars and other descriptive materials can attest. (They can be supplemented with problem-solving assignments that more directly relate to the concepts discussed in the textbook and

the readings.) Their value is to expose the students to the full richness and variety of human languages, which any linguistic theory tends to oversimplify in the name of creating order from data. If all goes well, this encounter engenders a fascination with ways of speaking (and perhaps of thinking) that are different from ours, and functions as an antidote to reductionist theorizing; and this is all for the best.

First in order of acknowledgment is the redwood country of the California Northcoast, in whose peaceful presence this book was largely written in the summer of 1988 (thanks to my family and to the University of Michigan, the latter through a Horace H. Rackham summer fellowship). Joseph Greenberg, Bernard Comrie, Keith Denning and three anonymous reviewers from the Cambridge Textbooks Series editorial board provided valuable comments on the earliest drafts. Special thanks go to Penny Carter of Cambridge University Press; Elizabeth Traugott of Stanford University; and Tom Toon of the University of Michigan. Pam Beddor exposed me to current work on phonological typology and phonetic explanation, some of which made its way into this volume. Four classes of typology students at Stanford and the University of Michigan contributed immeasurably to what ultimately became the organization of this volume. John Myhill read and commented on the penultimate draft, and used it in his typology course; Myhill's students gave important feedback on the manuscript, considerably improving the final version. Trisha Svaib assisted in preparing the final manuscript. Keith Denning provided valuable advice and invaluable moral support throughout the time that I wrote this volume.

Above all, I have benefited enormously from two of the leaders in the field of typology and universals. Bernard Comrie, whose research and whose own volume on typology set an excellent example for me, oversaw this project from the earliest drafts to the final manuscript and provided extensive comments and general support for my efforts. Finally, I must express my deeply felt appreciation to my teacher, Joseph H. Greenberg, whose erudition in human languages, language universals, historical linguistics and the history of linguistics is unequalled. I dedicate this volume to him with affection and respect.

Abbreviations

The abbreviations for grammatical morphemes and categories have been standardized in the examples, in accordance with the abbreviations adopted by the Framework for Descriptive Grammars project (Bernard Comrie, William Croft, Bruce Harold, Christian Lehmann and Dietmar Zaefferer) in 1991, and subsequently adopted (with some modifications) by the European Typology project. Abbreviations in this list have a maximum length of five, and were designed to eliminate ambiguity, maintain uniqueness of abbreviation, and to render some less-used abbreviations more 'natural'. All of the standard abbreviations are listed here, in the hope that their use will become more widespread. Some additional abbreviations found in the examples are also listed below. At the end of the list, abbreviations found in the text, where they are different from those found in the examples, are listed.

1	first person
2	second person
3	third person
12	first person dual inclusive (if treated as a quasi-singular)
A	transitive agent
ABESS	abessive ('without')
ABL	ablative ('from')
ABS	absolutive case
ABSL	absolute form
ABST	abstract (nominalization)
ACC	accusative
ACCID	accidental (action)
ACCESS	accessory (case)
ACT	active
ACTR	actor
ADESS	adessive ('on')
ADJR	adjectivalizer
ADM	admonitive
ADVRS	adversative
ADVR	adverbializer
AFF	affirmative
AFFCT	affective
AG	agent(ive)
AL	alienable
ALL	allative ('to')
ALLOC	allocutive
AN	animate
ANA	anaphoric
ANT	anterior
ANTI	antipassive
AOR	aorist
APPL	applicative
ART	article
ASP	aspect
ASS	assertive

ASSOC	associative
AT	attributor
AUG	augmentative
AUX	auxiliary
AVERS	aversive
BEN	benefactive
BUFF	phonological buffer element
CARD	cardinal (numeral)
CAUS	causative
CIRC	circumstantial
CLF	classifier
CLn	noun class n
CMPL	completive
CMPR	comparative
CONJ	conjunction
CJPRT	conjunctive participle
CO	co-ordinator
COLL	collective
COM	comitative
COMP	complementizer
CONC	concessive
COND	conditional
CONN	connective
CONST	construct form
CONT	continuous
CONTR	contrastive
COP	copula
CORR	correlative
CUST	customary
D1	deictic of 1 person
D2	deictic of 2 person
D3	deictic of 3 person
D12	deictic of 12 person
DAT	dative
DECL	declarative
DEF	definite
DEFR	deferential
DEM	demonstrative
DEP	dependent verb form
DER	derivational morpheme
DES	desiderative
DET	determiner
DETR	detransitivizer
DIM	diminutive
DIR	directional
DIST	distal (=3 person deictic)
DITR	ditransitive
DO	direct object
DS	different subject
DSTR	distributive
DU	dual
DUB	dubitative
DUR	durative
DWNT	downtoner
DYN	dynamic (vs. stative)
EL	elative ('out of')
EMPH	emphatic
EQT	equative (adjective)
ERG	ergative
ESS	essive ('as')
EVID	evidential
EX	exclusive
EXCL	exclamation
EXST	exist(ence)
F	feminine
FACT	factitive
FAM	familiar
FIN	finite
FNL	final position marker
FOC	focus
FREQ	frequentative
FRM	formal
FUT	future
G	ditransitive 'goal'
GEN	genitive
GER	gerund (verbal adverb)
GNR	generic
HAB	habitual
HEST	hesternal (past, future)
HOD	hodiernal (past, future)

HON	honorific
HORT	hortative
HUM	human
HYP	hypothetical
ILL	illative ('into')
IMM	immediate (past, future)
IMP	imperative
IMPF	imperfect(ive)
IMPR	impersonal
IN	inclusive
INAL	inalienable
INAN	inanimate
INCH	inchoative
INCP	inceptive
IND	indicative
INDF	indefinite
INESS	inessive ('in')
INF	infinitive
INFR	inferential evidential
INGR	ingressive
INJ	injunctive
INST	instrumental
INT	interrogative
INTR	intransitive
INTS	intensifier/intensive
INV	inverse
INVS	invisible
IO	indirect object
IRR	irrealis
ITER	iterative
JUSS	jussive
LIG	ligature
LNK	linker
LOC	locative
LOG	logophoric
M	masculine
MAL	malefactive
MAN	manner
MDL	modal
MED	medial (verb form)
MEDP	mediopassive
MEDT	mediate (= 2 person deictic)
MID	middle
MOD	modifier
N	neuter
NARR	narrative (tense)
NCLF	numeral classifier
NCMP	noncompletive
NEAR	near (past, future)
NEC	necessity
NEG	negative
NFNL	nonfinal position marker
NFOC	nonfocus
NFUT	nonfuture
NHUM	nonhuman
NOM	nominative
NPST	nonpast
NR	nominalizer
NSG	nonsingular
NSPEC	nonspecific
NTOP	nontopic
NVOL	nonvolitional
OBJ	object
OBL	oblique
OBLG	obligative
OBV	obviative
OPT	optative
ORD	ordinal (numeral)
P	transitive patient
PART	participle
PASS	passive
PAU	paucal
PCLF	possessive classifier
PEJ	pejorative
PFCT	perfect
PL	plural
PLT	pluritive
PLUP	pluperfect
PNCT	punctual
PO	primary object

POL	polite
POSS	possessive
POST	postposition
POT	potential
PRED	predicative
PREP	preposition
PREV	preverb
PRF	perfective
PRN	pronoun
PROG	progressive
PROH	prohibitive
PROL	prolative ('along')
PROX	proximal
PRS	present
PRT	preterit
PRTT	partitive
PRVT	privative ('without')
PRXT	proximate (= 1 person deictic)
PST	past
PTCL	particle
PURP	purpose, purposive
QUAD	quadral
QUOT	quotative
RDP	reduplication
REC	recent (past)
RECP	reciprocal
REF	referential
REFL	reflexive
REFR	referential ('about')
REL	relative clause marker (other than relative pronoun)
RPRN	relative pronoun
REM	remote (past, future)
REMT	remote (distance)
REP	reportive evidential
RES	resultative
RL	realis
RLT	relative (case)
S	intransitive subject
SBJ	subject
SENS	sensory evidential
SEQ	sequential, consecutive
SG	singular
SGT	singulative
SIM	simultaneous
SMLF	semelfactive
SO	secondary object
SPEC	specific
SS	same subject
STAT	stative
SUBJ	subjunctive
SUBR	subordinator
SUP	superlative
T	ditransitive 'theme'
TEMP	temporal
TERM	terminative
TNS	tense
TOP	topic
TR	transitive
TRNSF	transformative ('as')
TRNSL	translative ('becoming')
TRL	trial
TRNS	transitivizer
UNDR	undergoer
UNSP	unspecified (agent, etc.)
VAL	validator
VERS	version
VIS	visible
VISL	visual evidential
VN	verbal noun
VOC	vocative
VOL	volitional
VR	verbalizer

Additional abbreviations found in the text:

A, Adj	adjective
Adp	adposition
Adv	adverb
CN	common noun
G	genitive
L1	first language
N	noun
NP	noun phrase
Num	numeral
O	object
Ocmpr	object of comparison
Ocomp	object complement
Oprn	object pronoun
OT	Optimality Theory
PP	adpositional phrase
prn	pronoun
Purp	purpose clause
Q	interrogative particle
Rel	relative clause
S	subject
Sent	sentence
Std	standard of comparison
V	verb
VP	verb phrase

Symbols

The following symbols are used in example sentences in the original language and their interlinear morpheme translations, and in symbolic representations of syntactic structures. These symbols follow the conventions found in Lehmann (1982a), revised by the Framework for Descriptive Grammars project (Bernard Comrie, William Croft, Bruce Harold, Christian Lehmann and Dietmar Zaefferer) in 1991.

In both original language and interlinear morpheme translation:

x y	word boundary between x and y
x-y	morpheme boundary between x and y
x + y	x and y form a compound or a derivative stem
x = y	x and y are joined by clisis
$x_i \ldots y_i$	x and y are coreferential elements

In original language only:

Ø	null expression of meaning (optionally represented)
a<x>b	x is an infix, a . . . b is the discontinuous root/stem
a>y<b	a . . . b is a circumfix, y is the root/stem

In interlinear morpheme translation only:

(x)	x is not overtly marked in the original (i.e. null expression of meaning)
y<x>	x is the infix, a . . . b = y is the root/stem
<x>y	a . . . b = x is the circumfix, y is the root/stem
x \ y	y is an internal modification of lexeme x in the original
x:y	morpheme boundary between x and y not shown in the original
x.y	x and y are grammatical (sub)categories of one original language morpheme
x/y	x acts on y (indexation)
[x]	x is a syntactic constituent in the original language

$[x]_Y$	x is a syntactic constituent of category Y in the original language
[X Y Z]	a construction consisting of elements X, Y and Z, whose linear order is not necessarily fixed

The following logical symbols are used in the text in the formulation of language universals:

$P \& Q$	P and Q
$P \vee Q$	P or Q
$P \supset Q$	if P, then Q
$P \equiv Q$	P if and only if Q
$\sim P$	not P

1 Introduction

1.1 What is typology?

The term **typology** has a number of different uses, both within linguistics and without. The common definition of the term is roughly synonymous with 'taxonomy' or 'classification', a classification of the phenomenon under study into types, particularly structural types. This is the definition that is found outside of linguistics, for example in biology, a field that inspired linguistic theory in the nineteenth century.

The most unassuming linguistic definition of typology refers to a classification of structural types across languages. In this definition, a language is taken to belong to a single type, and a typology of languages is a definition of the types and an enumeration or classification of languages into those types. We will refer to this definition of typology as **typological classification**. The morphological typology of the nineteenth and early twentieth centuries is an example of this use of the term. This definition introduces the basic connotation that the term typology has in contemporary linguistics: typology has to do with **cross-linguistic comparison** of some sort. Methodological issues in cross-linguistic comparison will be discussed in §§1.3–1.6, while chapter 2 will be devoted to the notion of a linguistic type, including morphological typology, and its refinements in twentieth-century research.

A second linguistic definition of typology is the study of patterns that occur systematically across languages. We will refer to this definition of typology as **typological generalization**. The patterns found in typological generalization are language **universals**. The classic example of a typological universal is the implicational universal. An example of an implicational universal is the generalization, 'if the demonstrative follows the head noun, then the relative clause also follows the head noun.' This universal cannot be discovered or verified by observing only a single language, such as English. One has to do a general survey of languages to observe that the language type excluded by the implicational universal – namely a language in which the demonstrative follows the head noun and the relative clause precedes it – indeed does not exist.

Typological generalization is generally regarded as a subdiscipline of linguistics – not unlike, say, first language acquisition – with a particular domain of linguistic facts to examine: cross-linguistic patterns. Typology in this sense began in earnest with Joseph H. Greenberg's discovery of implicational universals of morphology and word order, first presented in 1960 (Greenberg 1966a). The primary purpose of the present volume is to discuss the kinds of cross-linguistic patterns that have been discovered and the methodological and empirical issues raised by the study of these patterns. Chapters 3–7 are devoted to discussing these patterns and the empirical and methodological issues that their discovery raises. The kinds of cross-linguistic patterns actually found represent a coherent set of language universals which are basic phenomena to be explained by any linguistic theory.

The third and final linguistic definition of typology is that typology represents an approach or theoretical framework to the study of language that contrasts with prior approaches, such as American structuralism and generative grammar. In this definition, typology is an approach to linguistic theorizing, or more precisely a methodology of linguistic analysis that gives rise to different kinds of linguistic theories than found in other approaches. Sometimes this view of typology is called the Greenbergian, as opposed to the Chomskyan, approach to linguistic theory (after their best known practitioners; see, for example, Smith 1982:256). This view of typology is closely allied to **functionalism**, the view that linguistic structure should be explained primarily in terms of linguistic function (the Chomskyan approach is contrastively titled **formalism**). For this reason, typology in this sense is often called the **(functional–)typological approach**, and will be called so here. More precisely, we may characterize this definition of typology as **functional–typological explanation**. The functional–typological approach became generally recognized in the 1970s; important figures beginning at that time include Givón, Haiman, Comrie, Hopper and Thompson. Functional–typological explanation has well-established historical antecedents, however (see Haiman 1985 and chapter 9), not least Greenberg himself.

The three linguistic definitions of typology correspond to the three stages of any empirical scientific analysis. Typological classification represents the observation of an empirical phenomenon (language) and classification of what we observe. Typological generalization – language universals – is the formation of generalizations over our observations. And the functional-typological approach constructs explanations of the generalizations over what we have observed. In this sense, typology represents an **empirical scientific** approach to the study of language.

Of course, in any empirical science the actual process of doing science does not proceed in these three discrete stages. In particular, explanations offer themselves at all stages in the scientific process. We will present typological explanations of

language universals as the universals themselves are introduced in chapters 3–7. The explanatory models used by typologists include competing motivations, economy, iconicity, processing, semantic maps in conceptual space, and a rethinking of syntactic argumentation. One significant dimension of typological explanation is that explanations of many grammatical phenomena are fundamentally diachronic, not synchronic. The diachronic approach requires a fundamental rethinking of typological principles, and is discussed in chapter 8. Chapter 9 then summarizes the approach to language that typology presents.

Not surprisingly, these differing definitions of typology – typological classification, typological generalization and functional–typological explanation/approach – have led to some confusion about what typology is, or is supposed to be. For example, it is sometimes claimed that typology is 'merely descriptive' or 'taxonomic'; that is to say, it does not provide a means for developing theories of language which can function as an alternative to, for example, generative linguistic theory. This represents a confusion of typological classification with typological generalization and explanation. Typological generalization represents a well-established method of analysis, and the typological approach is now a well-articulated approach to language.

The emphasis on theory and methodology in this volume should not be interpreted as minimizing the descriptive work necessary to develop typological analyses. The descriptive work which has been and, I hope, will continue to be done on the tremendous number of languages in the world is absolutely essential not just to typological theory but to all linguistic theories. Unfortunately, typological studies have often had to withhold or remove their data sections upon publication due to size limitations,[1] while many good descriptive works such as the University of Hawaii Press PALI series of Micronesian language grammars rapidly go out of print. The attitude that descriptive work is not valued (it is 'merely' descriptive or, disparagingly, 'descriptivist') must be abandoned for there to be progress in linguistic theory.

This matter becomes even more urgent because of the alarming loss of the empirical data base for linguistic theory. Hundreds of languages have become extinct in the last century. Hundreds, perhaps thousands, of others no longer survive in viable speech communities; the languages are dying and there are often serious consequences affecting grammatical structure. This situation is getting worse, not

[1] On some occasions, the data is published elsewhere. The data for Keenan and Comrie's study on the Noun Phrase Accessibility Hierarchy (Keenan and Comrie 1977; see chapter 5) was eventually published in another journal (Keenan and Comrie 1979); the data from Maxwell's study on linearization (Maxwell 1984) was published by a linguistics department (Maxwell 1985); and the data on Kortmann's study of adverbial subordinators in European languages (broadly construed; Kortmann 1997) was published on diskette by LINCOM Europa.

better, and is finally achieving the attention it deserves (Dorian 1981; Krauss 1992; Crystal 2000; Nettle and Romaine 2000). The empirical problems with language research parallel the problems in biological research, in particular in evolutionary theory and ecology: the extinction of languages and the loss of the linguistic communities is like the extinction of species and the loss of their habitat (ecosystems). In both disciplines it threatens theoretical progress.

1.2 Typology, universals and generative grammar

Greenberg's approach to language universals emerged at about the same time as Chomsky's, in the late 1950s. The conception of language universals in typology and generative grammar is quite different. In this section, we will briefly describe the emergence of Greenberg's and Chomsky's ideas, and the similarities and differences that are found in the two approaches to language (for more detailed discussion, see Hawkins 1988). We will return to the relationship between typology and generative grammar in later chapters in the context of more specific theoretical issues (see §§3.5, 7.2, 9.2–9.3).

Language universals reflect the belief that there exist linguistic properties beyond the essential definitional properties of language that hold for all languages. Although this belief has considerable modern currency, it is by no means a necessary fact or universally-held opinion, and in fact the opposite view was widely held until around 1960. To a considerable degree, the difference between the generative and typological approaches to language universals can be traced to the different traditions to which Chomsky and Greenberg responded. The generative approach represents a reaction against behavioristic psychology, while the typological approach represents a reaction against anthropological relativism.

The behaviorist view of language, in particular language learning, is anti-universalist in that it posits no innate, universal internal mental abilities or schemas. In the behaviorist view, linguistic competence is acquired through learning of stimulus–response patterns. In contrast, the generative approach posits the existence of innate internal linguistic abilities and constraints that play a major role in the acquisition of language. It is these constraints that represent linguistic universals in this approach. The argument used by Chomsky (e.g. Chomsky 1976) for the existence of innate universal linguistic competence refers to the 'poverty of the stimulus'. It is argued that the child has an extremely limited input stimulus, that is, the utterances that it is exposed to from the mother and other caregivers. This stimulus is incapable of permitting the child to construct the grammar of the adult's language in a classic behaviorist model; therefore, the child must bring innate universals of grammatical competence to bear on language acquisition. Hence,

the primary focus on universals in the generative tradition has been on their innate character.

The anthropological relativist view of language is that the languages of the world can vary arbitrarily: 'languages could differ from each other without limit and in unpredictable ways', in Martin Joos' famous passage (Joos 1957:96). This view of language was particularly strong among anthropological linguists studying North American Indian languages, which indeed differ radically in many ways from so-called Standard Average European languages. However, the comparison of one 'exotic' language or a limited number of languages to English only indicates diversity, not the range of variation, let alone limits thereto. Greenberg and others discovered that a more systematic sampling of a substantial number of languages reveals not only the range of variation but constraints on that variation. Those constraints demonstrate that languages do not vary infinitely, and the constraints represent linguistic universals. Hence, the primary focus on universals in the typological tradition has been on their cross-linguistic validity, and on universals that restrict possible language variation (see §3.1).

The innate universals posited by generative grammar are intended to explain linguistic structure. The poverty of the stimulus argument is essentially a deductive argument from first principles (although it does make assumptions about the nature of the empirical input, and what counts as relevant input). The poverty of the stimulus argument is one aspect of Chomsky's more generally **rationalist** approach to language. The universals posited by typology are intended to represent inductive generalizations across languages, in keeping with typology's **empiricist** approach to language. Typological universals call for explanation in terms of more general cognitive, social-interactional, processing, perceptual or other abilities. These abilities may also be innate, but they extend beyond language per se. The generative grammarian argues that the discovery of innate principles that the child brings to bear in learning a single language can be extrapolated to language in general (Chomsky 1981). The typologist argues that a grammatical analysis based on one language or a small number of languages will not suffice to reveal linguistic universals; only a systematic empirical survey can do so.

These differences in approach have led to claims that the Greenbergian approach and the Chomskyan approach to language universals and linguistic explanation are diametrically opposed to each other. In fact, there are significant similarities between the generative and (functional–)typological approaches. Both approaches begin with the analysis of language structure. Both approaches consider the central question of linguistics to be 'What is a possible human language?' (though see §§3.1, 8.1). Both approaches are universalist, in contrast to their predecessors. There is broad agreement that there do exist a substantial number of universals that hold of all languages (assuming attested exceptions can be accounted for by

other principled factors). For both approaches, the construction of linguistic generalizations involves abstraction over the data, though the Greenbergian abstracts patterns across languages and the Chomskyan abstracts patterns within languages (see §9.2). Likewise, explanations for linguistic universals rest on universal human abilities, which may or may not be language specific, and which probably have a significant innate component, though perhaps are not entirely innate. In fact, for both generative and typological approaches, the foundations of linguistic explanation are ultimately biological, although for the Chomskyan the biological basis is found in genetics (innate linguistic knowledge) and for the Greenbergian the biological basis is indirect, and is to be found in evolutionary theory (see §9.3; Croft 2000).

Nevertheless, there are two salient distinctive characteristics of the Greenbergian approach: the central role of cross-linguistic comparison, and the close relationship between linguistic form and language function. These two characteristics are discussed in the following two sections.

1.3 Cross-linguistic comparison

The first question that may be asked of typology is, what is the role of cross-linguistic comparison – the fundamental characteristic of typology – in linguistic analysis? Cross-linguistic comparison places the explanation of linguistic phenomena in a single language in a new and different perspective. For example, the distribution of the definite and indefinite articles in English is fairly complex:

(1a) He broke **a vase**.
(1b) He broke **the vase**.
(1c) The concert will be on **Saturday**.
(1d) He went to **the bank**.
(1e) I drank **wine**.
(1f) The French love **glory**.
(1g) He showed **extreme care**.
(1h) I love **artichokes** and asparagus.
(1i) Birds have **wings**.
(1j) His brother became **a soldier**.
(1k) **Dogs** were playing in the yard.

The eleven sentences given above characterize eleven types of uses of the articles (or their absence) in English, given as follows:

(a) specific (referential) indefinite (see §5.2);
(b) specific and definite;
(c) proper name;

(d) specific manifestation of an institution/place;
(e) partitive of a mass noun;
(f) generic mass noun;
(g) specific manifestation of an abstract quality (mass noun);
(h) generic of a count noun;
(i) generic of an indefinite number of a count noun;
(j) predicate nominal;
(k) specific but indefinite number of a count noun.

It might be possible to develop a set of generalizations – an **analysis** – that predicts exactly the distribution of the two articles (including their absence) in English. Such an account may be syntactic, semantic or pragmatic, or a combination of all three. Whatever is the case, it will have to be a fairly complex and subtle analysis, especially since the eleven different construction types given here do not exhaust the possibilities.

At this point, the typologist will ask: what is the significance of these generalizations posited in English for the class of human languages as a whole? Examining even a relatively closely related language, French, produces difficulties for those generalizations. In the exact same contexts, illustrated here by translation equivalents of the English sentences, the distribution of definite and indefinite articles *le/la/les* and *un/une* respectively (and their absence) is quite different:

(2a) Il a cassé **un vase**.
(2b) Il a cassé **le vase**.
(2c) Le concert sera **samedi**.
(2d) Il est allé à **la banque**.
(2e) J'ai bu **du vin**. (du = de + le)
(2f) Les Français aiment **la gloire**.
(2g) Il montra **un soin** extrême.
(2h) J'aime **les artichauts** et les asperges.
(2i) Les oiseaux ont **des ailes**. (des = de + les)
(2j) Son frère est devenu **soldat**.
(2k) **Des chiens** jouaient dans le jardin.

It is quite likely that the analysis of the distribution of the English articles would have to be drastically altered if not abandoned and a new one developed for the distribution of the French ones. In French, we find a more widespread use of both the French definite and indefinite articles, the appearance of the partitive marker *de* plus the definite article, and the absence of the French indefinite article in the predicate nominal construction.

One cannot be certain how much we would have to start all over again, of course, since to the best of my knowledge no complete analysis has been worked out. However, a generalization for a subset of three of the eleven contexts has been proposed,

for the generic count nouns in (h) and (i) and the indefinite number of count-noun usage in (k). Carlson (1977) proposes a unified analysis of the bare plural construction used in both situation types, in which both are of the same semantic type and the differing interpretations are attributed to the semantic type of the predicate. But when we turn to French, we see that in fact two different types of constructions are found – compare 2h and 2i,k – and so this generalization does not clearly apply to the grammatical facts of French. One may try to attribute the difference to the French partitive marker *de*. But if we turn to still other languages such as Rumanian (Farkas 1981:40–45), which distinguish the two uses solely by the presence vs. absence of the article, then we will not be able to invoke such an alternative.

The fact that analyses of linguistic phenomena 'one language at a time' cannot be carried over from one language to the next is somewhat disturbing for the search for language universals. Intricate interactions of internal structural generalizations are proposed by linguists to 'predict' grammatical patterns that do not apply even to neighboring languages. This is true not only in structuralist–generative analyses. Functionalist analyses, which invoke external (semantic or pragmatic) generalizations to account for the distribution of phenomena like the articles of English, often have the same problems:

> Volumes of so-called functionalism are filled with ingenious appeals to perception, cognition or other system-external functional domains, which are used to 'explain' why the language in question simply has to have a grammatical particularity that it does – when a moment's further reflection would show that another well-known language, or even just the next dialect down the road, has a grammatical structure diametrically opposed in the relevant parameter. (DuBois 1985:353)

The question here is, to what level of generalization should an analysis of language-specific facts be developed before taking into consideration cross-linguistic patterns? The typologist essentially takes the position that cross-linguistic patterns should be taken into consideration at virtually every level of generalization about human languages (see §9.3).

A cross-linguistic comparative approach – that is the construction of typological generalizations – allows us to make progress on universal characteristics of the distribution of articles, for example, and in turn causes us to reassess an analysis formulated without reference to the facts in other languages.

There are certain generalizations that cut *across* the two languages that are very likely to be characteristic of language in general. For instance, the first three uses, (a)–(c), are identical in English and French, and it is only in the following seven that there is substantial variation between the two languages. With the exception of the (k) use, all of the variable uses across the two languages concern generic

and mass-noun contexts of various sorts. This suggests that there may be some degree of uniformity across languages in specific NP contexts that does not exist in generic and mass NP contexts. (In fact, there is also variation in specific NP contexts, but of a more constrained type; see §8.2.)

There are two important points implicit in this proposed generalization over the English and French facts which summarize the argument for cross-linguistic comparison. The first is that this generalization could not be formulated without looking at more than one language. (Examining still more languages would, of course, further refine this generalization.) That is what makes this analysis of the grammatical phenomenon typological.

The second point pertains to the description and analysis of the grammar of a particular language, given the sorts of cross-linguistic generalizations that exist. Awareness of cross-linguistic variation allows the linguist describing a particular language to provide a more fine-grained description of the phenomenon in question. For example, being aware of the differences between English and French in generic and mass-noun contexts implies that a grammatical description should explicitly indicate how a language with articles expresses or **encodes** those different semantic types of NPs.

A fine-grained description of the linguistic facts of a language is sufficient for descriptive completeness. Of course, one always wants to seek generalizations in the data. Moreover, one would like the generalizations to correspond to some empirically real phenomenon, such as a speaker's knowledge of her (or his) language. If the generalizations are intended to represent a speaker's knowledge of her language, then such an analysis must integrate cross-linguistic comparison, according to the typological approach. For example, the generalizations about the distribution of the articles in both English and French ought to characterize the distribution in specific NP contexts in each language as typical or even universal (if that turns out to be the case), and the distribution in generic and mass NP contexts as arbitrary and language specific, or perhaps subject to other conditions that would be revealed by further cross-linguistic comparison. In this view, the analysis of the articles in French or English would be incomplete – and therefore an inadequate explanation of the phenomenon – if its relationship to cross-linguistic generalizations about articles is not taken into account. The generalizations revealed by examining more than one language at a time are the only ones which can be said to hold of languages in general. A speaker's knowledge of her language involves both universal and language-particular properties.

Until relatively recently, typology has not directed its attention to the relationship between language universals and the generalizations posited in particular language grammars (Croft 1999; §9.1). However, it is not the case that language universals exist independently apart from the linguistic knowledge of language