

OXFORD
LINGUISTICS

language

universals

diversity

unity

Linguistic Typology

Jae Jung Song

categories

OXFORD TEXTBOOKS IN LINGUISTICS

Linguistic Typology

OXFORD TEXTBOOKS IN LINGUISTICS

PUBLISHED

The Grammar of Words
An Introduction to Linguistic Morphology
THIRD EDITION
by Geert Booij

A Practical Introduction to Phonetics
SECOND EDITION
by J. C. Catford

An Introduction to Multilingualism
Language in a Changing World
by Florian Coulmas

Meaning in Use
*An Introduction to Semantics
and Pragmatics*
THIRD EDITION
by Alan Cruse

Natural Language Syntax
by Peter W. Culicover

Principles and Parameters
An Introduction to Syntactic Theory
by Peter W. Culicover

A Semantic Approach to English
Grammar
by R. M. W. Dixon

Semantic Analysis
A Practical Introduction
by Cliff Goddard

Pragmatics

SECOND EDITION
by Yan Huang

Compositional Semantics
*An Introduction to the
Syntax/Semantics Interface*
by Pauline Jacobson

The History of Languages
An Introduction
by Tore Janson

The Lexicon
An Introduction
by Elisabetta Ježek

A Functional Discourse
Grammar for English
by Evelien Keizer

Diachronic Syntax
by Ian Roberts

Linguistic Typology
by Jae Jung Song

Cognitive Grammar
An Introduction
by John R. Taylor

Linguistic Categorization
THIRD EDITION
by John R. Taylor

IN PREPARATION

Translation
Theory and Practice
by Kirsten Malmkjaer

Grammaticalization
by Heiko Narrog and Bernd Heine

Speech Acts and Sentence Types in English
by Peter Siemund

Cognitive Grammar
An Introduction
SECOND EDITION
by John R. Taylor

Linguistic Typology

JAE JUNG SONG

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© Jae Jung Song 2018

The moral rights of the author have been asserted

First Edition published in 2018

All rights reserved. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted, in any form or by any means, without the
prior permission in writing of Oxford University Press, or as expressly permitted
by law, by licence or under terms agreed with the appropriate reprographics
rights organization. Enquiries concerning reproduction outside the scope of the
above should be sent to the Rights Department, Oxford University Press, at the
address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2017939053

ISBN 978-0-19-967709-2 (hbk.)

978-0-19-967749-8 (pbk.)

Printed and bound by
CPI Group (UK) Ltd, Croydon, CRO 4YY

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

Contents

<i>Foreword</i>	xiii
<i>Preface</i>	xv
<i>List of abbreviations</i>	xvii
<i>Genealogical affiliations and geographical locations of languages cited in the book</i>	xix
1. Linguistic typology: An introductory overview	1
1.1 Introduction	1
1.2 What is linguistic typology?	1
1.3 Linguistic typology: a short history	5
1.4 Concluding remarks	13
<i>Study questions</i>	13
<i>Further reading</i>	14
PART I Foundations: theory and method	15
2. Unity and diversity in the world's languages	17
2.1 Introduction	17
2.2 The connection between diversity and unity	20
2.3 The Principle of Uniformitarianism: a methodological frame of reference	22
2.4 When and where similarities count	26
2.5 Types of language universals and universal preferences	30
2.6 Concluding remarks	36
<i>Study questions</i>	37
<i>Further reading</i>	38
3. Typological analysis	39
3.1 Introduction	39
3.2 'Comparing apples with apples': cross-linguistic comparability	44

CONTENTS

3.3	Comparative concepts vs descriptive categories	51
3.4	Concluding remarks	59
	<i>Study questions</i>	60
	<i>Further reading</i>	61
4.	Linguistic typology and other theoretical approaches	62
4.1	Introduction	62
4.2	Conceptual differences between LT and GG	64
4.3	Methodological differences between LT and GG	68
4.4	Optimality Theory: a derivative of GG with a twist	71
4.5	Concluding remarks	74
	<i>Study questions</i>	75
	<i>Further reading</i>	76
5.	Language samples and sampling methods	77
5.1	Introduction	77
5.2	Some recalcitrant issues in language sampling	79
5.3	Types of language sample	83
5.4	Biases in language sampling and how to avoid them	88
5.5	Independence of cases	92
5.6	Sampling procedures	94
5.6.1	Proportional representation in language sampling	94
5.6.2	Independence of cases in proportional representation	97
5.6.3	Having the best of both worlds: structural variation across and within phyla	99
5.7	Testing independence of cases at a non-genealogical level	102
5.8	Typological distribution over time: was it different then from what it is now?	110
5.9	Concluding remarks	115
	<i>Study questions</i>	116
	<i>Further reading</i>	117
6.	Data collection: Sources, issues, and problems	118
6.1	Introduction	118

CONTENTS

6.2 Grammatical descriptions or grammars	119
6.3 Texts	122
6.4 Online typological databases	126
6.5 Native speaker elicitation: direct or indirect	129
6.6 Levels of measurement and coding	134
6.7 Concluding remarks	137
<i>Study questions</i>	137
<i>Further reading</i>	138
7. Typological asymmetry: Economy, iconicity, and frequency	140
7.1 Introduction	140
7.2 Typological asymmetry	144
7.2.1 Formal coding	144
7.2.2 Grammatical behaviour	149
7.3 Economy and iconicity (in competition)	154
7.4 Typological asymmetry = frequency asymmetry?: iconicity vs frequency	158
7.5 Concluding remarks	167
<i>Study questions</i>	168
<i>Further reading</i>	169
8. Categories, prototypes, and semantic maps	170
8.1 Introduction	170
8.2 Classical vs prototype approach to categorization	171
8.3 Prototype category as a network of similarities	174
8.4 Prototype in grammar	176
8.5 Semantic maps: 'geography of grammatical meaning'	187
8.6 Concluding remarks	196
<i>Study questions</i>	197
<i>Further reading</i>	198

CONTENTS

PART II	<i>Empirical dimensions</i>	199
9.	Phonological typology	201
9.1	Introduction	201
9.2	Segmental typology	204
9.2.1	Consonants	204
9.2.2	Vowels	209
9.2.3	Consonant–vowel ratios	212
9.3	Syllabic typology	213
9.4	Prosodic typology	219
9.4.1	Tone	220
9.4.2	Stress	222
9.5	Concluding remarks	224
	<i>Study questions</i>	225
	<i>Further reading</i>	225
10.	Basic word order	227
10.1	Introduction	227
10.2	Some basic word order patterns	229
10.3	Early word order research	231
10.3.1	Greenberg’s seminal work and other ‘derivative’ works	231
10.3.2	Bringing word order inconsistencies to order	235
10.3.3	Distribution of the six basic clausal word orders	238
10.4	The OV–VO typology and the Branching Direction Theory	243
10.4.1	Back to the OV–VO typology	243
10.4.2	Inadequacy of Head-Dependent Theory	245
10.4.3	Branching Direction Theory (BDT)	249
10.4.4	Further thoughts on BDT	253
10.5	Word order variation and processing efficiency	255
10.5.1	Basic assumptions of EIC Theory	255
10.5.2	Left–right asymmetry in word order	263
10.6	Structural complexity and efficiency	266
10.6.1	Processing principles and processing domains	267

CONTENTS

10.6.2 Word order and word order correlations in Hawkins's extended theory	270
10.7 Areal word order typology	274
10.7.1 Areal distribution of the six clausal word orders	276
10.7.2 Areal distribution of OV and VO	277
10.7.3 Areal distribution of OV/VO and NPo/PrN	277
10.7.4 Areal distribution of OV/VO and RelN/NRel	278
10.8 Concluding remarks	279
<i>Study questions</i>	280
<i>Further reading</i>	281
11. Case alignment	282
11.1 Introduction	282
11.2 S-alignment types	288
11.2.1 Nominative–accusative type	288
11.2.2 Ergative–absolutive type	290
11.2.3 Tripartite type	292
11.2.4 Double oblique type	293
11.2.5 Neutral type	294
11.3 Variations on S-alignment	294
11.3.1 Split-ergative type	295
11.3.2 Active–stative type	298
11.3.3 Hierarchical type	302
11.4 Distribution of the S-alignment types	303
11.5 Explaining various case alignment types	306
11.5.1 The discriminatory view of case marking	306
11.5.2 The indexing view of case marking	309
11.5.3 The discriminatory view vs the indexing view	311
11.6 The Nominal Hierarchy and the split-ergative system	312
11.7 Case marking as an interaction between attention flow and viewpoint	316
11.8 P-alignment types	323
11.9 Distribution of the P-alignment types	326
11.10 Variations on P-alignment	329
11.11 S-alignment and P-alignment in combination	330

CONTENTS

11.12 Case alignment and word order	331
11.13 Concluding remarks	336
<i>Study questions</i>	336
<i>Further reading</i>	338
12. Grammatical relations	339
12.1 Introduction	339
12.2 Agreement	348
12.3 Relativization	353
12.4 Noun phrase ellipsis under coreference	359
12.5 Hierarchical nature of grammatical relations	364
12.6 Concluding remarks	367
<i>Study questions</i>	367
<i>Further reading</i>	370
13. Valency-manipulating operations	371
13.1 Introduction	371
13.2 Change of argument structure with change of valency	372
13.2.1 Passive	373
13.2.2 Antipassive	378
13.2.3 Noun incorporation	381
13.2.4 Applicative	383
13.2.5 Causative	391
13.3 Change of argument structure without change of valency	393
13.4 Concluding remarks	398
<i>Study questions</i>	399
<i>Further reading</i>	401
14. Person marking	403
14.1 Introduction	403
14.2 Morphological form: variation and distribution	411
14.2.1 Person marking and case alignment	412
14.2.2 Person marking and grammatical relations	414

CONTENTS

14.3 Paradigmatic structure: towards a typology	417
14.3.1 Grouping A: no inclusive/exclusive with split group marking	425
14.3.2 Grouping B: no inclusive/exclusive with homophonous group marking	428
14.3.3 Grouping C: inclusive/exclusive with split group marking	430
14.3.4 Grouping D: inclusive/exclusive with homophonous group marking	432
14.3.5 Structural dependencies in paradigmatic structure	433
14.4 Concluding remarks	437
<i>Study questions</i>	437
<i>Further reading</i>	439
15. Evidentiality marking	440
15.1 Introduction	440
15.2 Morphological form of evidentiality marking	442
15.3 Semantic parameters of evidentiality	444
15.3.1 Visual	444
15.3.2 Non-visual sensory	446
15.3.3 Inference and assumption	447
15.3.4 Hearsay and quotative	449
15.3.5 Order of preference in evidentials	451
15.4 Typology of evidentiality systems	452
15.4.1 Two-term systems	452
15.4.2 Three-term systems	454
15.4.3 Four-term systems	457
15.4.4 Languages with more than four evidentiality distinctions	459
15.4.5 Multiple evidentiality subsystems	460
15.5 Evidentiality and other grammatical categories	460

CONTENTS

15.6 Concluding remarks	462
<i>Study questions</i>	463
<i>Further reading</i>	464
<i>References</i>	465
<i>Author index</i>	485
<i>Language index</i>	489
<i>Subject index</i>	496

Foreword

Jae Jung Song died peacefully at home in Dunedin, New Zealand on 24 April 2017 while this book was being prepared for publication. Born on 18 January 1958 to Tae Hyun Song and Jin Sun Yoo, Jae was one of four siblings, including his brother, Jae Tag. He is survived by his son, Kee, daughter-in-law Coral, and his grandson, whose birth Jae was eagerly anticipating, but sadly missed.

Jae received a BA Honours and PhD from Monash University, Melbourne, Australia. After several years teaching in Australia and Singapore, Jae joined the University of Otago in 1992. Jae was instrumental in re-establishing and expanding Otago's Linguistics Programme to include a TESOL minor and a Graduate Diploma for Second Language Teachers. During his career he served on the editorial boards of the *Web Journal of Formal, Computational and Cognitive Linguistics*, the *Australian Journal of Linguistics*, the *International Review of Korean Studies*, *Language and Linguistics Compass*, and the *New Zealand Journal of Asian Studies*. He was also Associate Editor of the journal *Linguistic Typology*. This book is Jae's tenth authored or edited book, capping a career that also included publication of thirty-two book chapters and forty journal articles. Although Jae is most widely known for his extensive contributions to linguistic typology, particularly in the area of causatives, he also published on South and North Korean language policy.

Many of Jae's colleagues are unaware that Jae was a survivor of polio, which he contracted as a child in South Korea. At the height of his career he was struck by post-polio syndrome, a virtually untreatable re-emergence of the effects of polio. In the ten years I worked with Jae, he progressed through the support of a cane, one crutch, two crutches, a mobility chair and crutches, and finally a motorized wheelchair. Jae was determined not to allow the syndrome to compromise his work, and maintained a usual workload alongside his impressive publication record.

FOREWORD

Jae will be remembered by colleagues, students, and friends for his work in typology, but he will be missed for his unwavering determination, acute intelligence, and ready wit.

Anne Feryok
University of Otago
Department of English and Linguistics
Dunedin, New Zealand

Preface

Not many authors are fortunate enough to have an opportunity to write more than one introductory volume in the same field. I am one of the fortunate ones. My first introductory book, *Linguistic Typology: Morphology and Syntax* (Harlow, 2001), appeared sixteen years ago, and it was in great need of updating, addition, and revision, in view of a substantial amount of developments—theoretical and methodological—and new data drawn from previously as well as newly documented languages. Two introductory books have recently appeared, namely Viveka Velupillai's *An Introduction to Linguistic Typology* (Amsterdam, 2012) and Edith Moravcsik's *Introducing Language Typology* (Cambridge, 2013). These two are excellent introductions to the field and continue to inspire and inform students and professional linguists alike. They are, however, written primarily with students with minimal prior knowledge of linguistics in mind, and there is a need for a text that is pitched at an advanced level. I envisage that readers at this higher level of study are capable of, and keen on, grappling with theoretical or methodological issues that have characterized most of the recent developments in linguistic typology. I hope that my new book will meet that need. My earlier book was organized largely by topic area, which makes it different from William Croft's *Typology and Universals* (Cambridge, 2003), which is organized by theoretical concept. The present book aims to strike a balance between these two different styles of organization, which I believe is the best way to introduce my target audience to the field.

I would like to thank John Davey (the former Commissioning Editor at Oxford University Press) for inviting me to write a proposal for this book, and his successor Julia Steer for her forbearance and support, especially when the writing of the book ran into difficulty, more than once, because of my personal circumstances. I would also like to mention a number of linguists whose writings have influenced, or contributed to, my thinking not only about the various issues addressed in this book but also about many others. There are too many to list them here but the following eminent scholars deserve a special mention: Balthasar

PREFACE

Bickel, Barry Blake, Guglielmo Cinque, Bernard Comrie, William Croft, Matthew Dryer, the late Joseph Greenberg, John Haiman, Martin Haspelmath, John Hawkins, Edith Moravcsik, Johanna Nichols, and, last but not least, the late Anna Siewierska. The writing of this book has also benefited from the invaluable input of students in the typology course that I have taught for the last twenty-five years at the University of Otago. One of the advantages of teaching, as many teachers will agree, is that one always tries to find a better way to explain things to students, and, hopefully, this advantage has made its way into the book. There are also a number of people who have contributed non-linguistically to the birth of this book. I am indebted to Jaetag Song and Kee Ho Song for their words of encouragement. I cannot forget to thank Eleni Witehira for her unfailing support, even in times of her own personal difficulties, and also Kun Yong Lee and Sunae Bang (Sushi Station, Dunedin) for often cooking something off the menu for me, despite their heavy workload. I also wish to express my gratitude to the following people who have never heard about linguistic typology and have kindly provided me with non-work-related assistance and support, without which the writing of this book would have been near impossible: Sarah Andrews, Lisa Begg, Dr Rene Cescon, Dr Martin Dvorcek, Hannah Fleming, Linda Grady, Lynda Hurren, Toni Johnston, Sandi Lorincz, Miriam Mackay, Michelle Mielnik, Zena Pigden, Ange Price, Dr Markus Renner, Nic Rogan, and Tammy Waugh. Lastly, I would like to thank Lisa Marr for her assistance in checking the references and preparing the indices.

List of abbreviations

1	first person	DIR.EV	direct evidential
2	second person	DISTR	distributive
3	third person	DO	direct object
ABL	ablative	DR	direct
ABS	absolutive	DU	dual
ACC	accusative	EMPH	emphasis
ALL	allative	EP	epenthetic
ANIM	animate		formative
ANTIP	antipassive	ERG	ergative
AOR	ao­rist	EXCL	exclusive
APPL	applicative	EZ	ezāfe
ART	article	F	feminine
ASP	aspect	FIRST	first-hand
ASSUM	assumed	FOC	focus
ATTRIB	attributive	FUT	future
AUX	auxiliary	GEN	genitive
BEN	benefactive	IMM	immediate
CAUS	causative	IMP	imperative
CLF	classifier	INAN	inanimate
CLT	clitic	INCH	inchoative
CNTR	contrast	INCL	inclusive
COM	comitative	INCOMP	incomplete
COMP	complementizer	IND	indicative
COMPL	completive	INDEF	indefinite
COP	copula	INF	inferred
DAT	dative	INFV	infinitive
DEC	declarative	INST	instrumental
DEF	definite	INT	intentional
DEM	demonstrative	INTR	intransitive
DET	determiner		marker

LIST OF ABBREVIATIONS

INTRATERM	intraterminal	PL	plural
INV	inverse	POSS	possessive
IO	indirect object	PREP	preposition
IPFV	imperfective	PRES	present
IRLS	irrealis	PROCOMP	procomplement
LINK	linker	PRON	pronoun
LOC	locative	PROX	proximate
M	masculine	PRTV	partitive
N	neuter	PST	past
NARR	narrative	PV	preverb
NCL	noun class	Q	question
NEG	negative	QUOT	quotative
NF	non-final marker	REC.P	recent past
NFUT	non-future	REF	referential
NOM	nominative	REFL	reflexive
NOMN	nominalization/ nominalizer	REL	relative
		REP	reportative
NON.FIRST	non-first-hand	RETRO	retrospective
NONVIS	non-visual	SBJ	subject
NR.PST	near past	SBJV	subjunctive
NUM	number	SD	sudden discovery
OBJ/O	object	SENS.EV	sensory evidential
OBL	oblique	SEQ	sequence
OBV	obviative	SG	singular
OPT	optative	SUB	subordinating
PART	participle	SUBR	subordinator
PASS	passive	TES	testimonial
PAUC	paucal	TNS	tense
PC	particle of concord	TOP	topic
		TR	transitive marker
PEJ	pejorative	TRANSLOC	translocative
PERF	perfect	V	verb
PF	phonological filler	VIS	visual
PFV	perfective	VOL	volitional

Genealogical affiliations and geographical locations of languages cited in the book

Each language name is followed by (X; Y: Z), where X is the name of its genus, Y the name of its language family, and Z the name of the country where it is spoken, e.g. Bayso (Eastern Cushitic; Afro-Asiatic: Ethiopia). Languages that do not belong to any known language family are identified as ‘isolates’, e.g. Korean (isolate: Korea). If there is no distinction between genus and family (i.e. a single-genus language family) or between language and genus (i.e. a single-language genus), only one genealogical name is given, e.g. Pirahã (Muran: Brazil) or Palauan (Austronesian: Palau). Genealogical and geographical information is omitted where it is deemed redundant, that is, given elsewhere in the relevant section or chapter. For further genealogical and geographical information on the world’s languages, the reader may like to visit:

- (1) <http://wals.info/languoid>;
- (2) <http://glottolog.org/glottolog/language>; and
- (3) <https://www.ethnologue.com/browse/names>.

1

Linguistic typology

An introductory overview

1.1 Introduction	1
1.2 What is linguistic typology?	1
1.3 Linguistic typology: a short history	5
1.4 Concluding remarks	13

1.1 Introduction

This chapter provides a brief description of linguistic typology by explaining what it is, what it aims to achieve in its study of language, what kinds of question it raises about the nature of language, and how it does what it does. The description will give the reader a snapshot of linguistic typology before they embark on reading the rest of this book. This is followed by a historical overview of linguistic typology. It is important to learn how linguistic typology emerged and developed over time into what it is today, as the reader will need to have a background understanding of how linguistic typology has evolved conceptually into one of the most important theoretical approaches to the study of language.

1.2 What is linguistic typology?

The reader may have come across the kind of restaurant menu to be described here (more commonly found in Chinese than other restaurants): à la carte menu items listed under different main-ingredient

headings, e.g. beef, pork, chicken, duck, seafood, and vegetables. For instance, all the dishes containing beef as their main ingredient are grouped together under the heading of ‘beef’; the dishes with seafood as their main ingredient are all listed together under the heading of ‘seafood’, and so on. In other words, à la carte dishes are put into different types according to the main ingredients used. The basic principle employed in this way of organizing the menu is typological in that the main ingredient (e.g. beef) is used as the criterion for classifying or typologizing a given dish (e.g. beef with ginger and spring onion) in a particular way (e.g. a type of beef dish). This menu can thus be thought of as a typology of the dishes contained therein. The dishes may involve secondary ingredients as well, and any one of these ingredients may alternatively be selected as the criterion for classifying the dishes (e.g. noodle dishes vs rice dishes). Different ingredients (i.e. typological properties), when chosen as the basis for typologizing, may give rise to different typologies of the same dishes. To wit, typology (e.g. a particular way of organizing the menu items) is the classification of a domain (e.g. the menu).

The same principle can be applied to the study of languages. Thus, the world’s languages can be put into different types according to a particular linguistic property. The output of this exercise will be a typology of languages or, specifically, a linguistic typology of languages, since the typological property used is a linguistic one. For instance, take adpositions, which are linguistic elements expressing the semantic relations between the verb and its related noun phrases, e.g. location, time, instrument. The English adposition *in* in (1) expresses the location of Rachel’s action: Rachel’s action took place *in*, as opposed to *near*, the garden.

- (1) English (Germanic; Indo-European: UK)
Rachel washed her car *in* the garden.

The adposition in (1) is known as a preposition, because it appears right before the noun phrase that it associates itself with. In languages such as Korean, adpositions (e.g. *-eyse* ‘in’) are placed after the noun phrase, as in (2), in which case they are known as postpositions:

- (2) Korean (isolate: Korea)
kiho-ka anpang-eyse thipi-lul po-ass-ta
Keeho-NOM main.bedroom-LOC TV-ACC see-PST-IND
‘Keeho watched TV in the main bedroom.’

Using the two types of adposition, it is now possible to classify languages into two different types: prepositional and postpositional languages.¹ For instance, English is identified as a prepositional language, and Korean as a postpositional language. The use of adpositions as the basis of typologizing languages is a simple illustration of how to put languages into different types according to a particular typological property (or, simply put, how to typologize languages). Other typological properties may be selected as the basis for classifying languages. For instance, the grammatical roles within the clause, e.g. subject and object, may be expressed on the head or the non-head (aka the dependent) (for detailed discussion, see Chapter 11). Note that in the clause, the verb is the head and its arguments (or the noun phrases that the verb associates itself with) are dependents. In Swahili, for example, the grammatical roles of the noun phrases are all indicated on the verb.

- (3) Swahili (Bantoid; Niger-Congo: Tanzania)
 Ahmed a-li-m-piga Badru
 Ahmed he-PST-him-hit Badru
 ‘Ahmed hit Badru.’

In (3), the verb (or the head) contains overt marking (i.e. *a-* for the subject noun phrase and *m-* for the object noun phrase). In contrast, the noun phrases (or the dependents), *Ahmed* and *Badru*, do not have any marking that indicates their grammatical roles. This type of marking is known as head-marking (i.e. the head is marked). In Pitta-Pitta, the overt marking of the grammatical roles of the noun phrases in the clause appears on the dependents (= the noun phrases) themselves, as in:

- (4) Pitta-Pitta (Central Pama-Nyungan; Pama-Nyungan: Australia)
 kaŋa-lu matjumba-ŋa piŋi-ka
 man-ERG roo-ACC kill-PST
 ‘The man killed the kangaroo.’

Not unexpectedly, the marking type exemplified in (4) is referred to as dependent marking (i.e. the dependents are marked). Thus, according

¹ Less commonly found are languages with both prepositions and postpositions, and even much less commonly attested are languages with inpositions, which appear inside the noun phrase. For details on inpositions, see Dryer (2013c). For the sake of simplicity, these languages will be ignored here.

to the locus of grammatical-role marking in the clause, languages are classified into the head-marking or dependent-marking type (e.g. Nichols and Bickel 2013a). Swahili is a head-marking language, whereas Pitta-Pitta is a dependent-marking language.²

While linguistic typology shares the same conceptual basis of classification with the Chinese menu described at the beginning of this section, that's where their similarity ends. Unlike the menu, linguistic typology goes well beyond the business of classifying its objects of inquiry (i.e. languages). In fact, the more fruitful part of linguistic typology begins only after the establishment of a typological classification of languages. For concreteness, take adpositions again. Broadly speaking, the world's languages can be classified into either the prepositional or the postpositional type—note that, if a language employs both prepositions and postpositions, it may still be possible to say that the language is either predominantly prepositional or predominantly postpositional. If the investigation stops at this point, doing linguistic typology will be rather bland, if not boring, although that investigation in itself may be a rewarding experience. Still, it is hardly the most exciting thing to point to the mere existence of prepositional languages and postpositional languages in the world without saying any more. This is why Joseph Greenberg (1915–2001), regarded as the father of modern linguistic typology, wrote: 'The assignment of a language to a particular typological class becomes merely an incidental by-product and is not of great interest for its own sake' (1978a: 40). So, what can be done to make it intellectually stimulating and challenging? This is what needs to be done. One must make every effort to analyse or interpret a given typological investigation; one must attempt to account for what the typological investigation has produced, a linguistic typology of X, with a view to finding out about the nature of human language. To wit, one must make sense of the 'incidental by-product'. But how does one make sense of the linguistic typology of adpositions, for instance? This particular question can be formulated in such a way as to ascertain what factors motivate the use of prepositions or postpositions. Asked differently, do languages choose randomly or systematically between prepositions and postpositions? This question, in turn, may lead to

² For the sake of simplicity, languages with both head- and dependent-marking or with neither or with marking appearing elsewhere in the clause will be ignored. For details, see Nichols (1986), and Nichols and Bickel (2012a, 2012b, 2012c).

further investigation of adpositions. Indeed, linguistic typologists have discovered that the use of prepositions or postpositions is not an independent phenomenon. There is a remarkably strong correlation between the types of adpositions and basic clausal word order (specifically the position of the verb in the clause) to the extent that one can confidently state that verb-initial languages (i.e. verb appearing before subject and object in the clause) have prepositions, and verb-final languages (i.e. verb appearing after subject and object in the clause) postpositions.³ In other words, what seem to be two logically independent properties (i.e. adpositions and basic word order) correlate with each other to the point of statistical significance (for further discussion, see Chapter 10). Imagine that linguistic typologists do not venture beyond the mere classification of languages into the two adpositional types—that is, beyond the level of an incidental by-product. They will not be able to discover the correlations in question, let alone to find themselves in a position to pose a far more intriguing, albeit challenging, question: why do these correlations exist in the first place? That is, what is it that makes verb-initial and verb-final languages opt for prepositions and postpositions, respectively? Other related questions may include: what is so special about the position of the verb that has a bearing on the choice between prepositions and postpositions? Is there something shared by the position of the verb and the type of adposition that it correlates with? Asking questions such as these is important because they may lead to an understanding of the nature of language. Furthermore, such an understanding may provide useful insights into the way the human mind shapes language or, more generally, how the human mind works (see Chapter 10).

1.3 Linguistic typology: a short history

The prospect of being able to predict the presence of one property (e.g. the types of adposition) on the basis of the presence of another property (e.g. the position of the verb in the clause) is a very attractive one. First and foremost, this elevates linguistic typology to a different level of investigation: so-called implicational typology. Even better will be the

³ Verb-initial and verb-final order have subsequently been generalized to VO and OV order, respectively. Thus, VO order prefers prepositions and OV order postpositions.

presence of one property (X) implying the presence of multiple properties (Y, Z, etc.). For instance, there may be properties other than the use of prepositions or postpositions that verb-initial or verb-final order correlates with. Indeed, verb-initial order is also known to prefer Noun–Genitive order (e.g. *the residence of the Prime Minister*) to Genitive–Noun order (e.g. *the Prime Minister’s residence*), whereas verb-final order also has a strong tendency to co-occur with Genitive–Noun order (for further discussion, see Chapter 10). Generally speaking, it is preferable to predict as much as possible on the basis of as little as possible. This situation is akin to generating multiple units of energy (e.g. electricity) by using one unit of energy (e.g. biofuel). One cannot fail to see an economic dimension at work here.

The idea of using the presence of one property to draw inferences about that of other properties is not something that has recently come to the attention of linguistic typologists. It has actually been around for a very long time. Over a hundred years ago, the German scholar, Georg von der Gabelentz (1840–93), was so excited about this idea that he expressed his thinking somewhat overenthusiastically:

Aber welcher Gewinn wäre es auch, wenn wir einer Sprache auf den Kopf zusagen dürften: Du hast das und das Einzelmerkmal, folglich hast du die und die weiteren Eigenschaften und den und den Gesamtcharakter!—wenn wir, wie es kühne Botaniker wohl versucht haben, aus dem Lindenblatte den Lindenbaum construiren könnten. (Gabelentz 1901: 481)

[But what an achievement would it be were we to be able to confront a language and say to it: ‘you have such and such a specific property and hence also such and such further properties and such and such an overall character’—were we able, as daring botanists have indeed tried, to construct the entire lime tree from its leaf. (translation by Shibatani and Bynon 1995b: 10)]

This statement, often cited in textbooks on linguistic typology, is probably the earliest articulation of linguistic typology, as is currently understood and practised. Indeed, ‘[i]t would be difficult to formulate the research programme of linguistic typology more succinctly’ than Gabelentz did (Plank 1991: 421). Moreover, it was Gabelentz (1901: 481) who coined the term ‘typology’ when he continued to write on the same page: ‘Dürfte man ein ungeborenes Kind taufen, ich würde den Namen Typologie wählen’ [If one were permitted to christen an unborn child, I would choose the name typology]. The term ‘typology’, usually in conjunction with a modifying expression ‘linguistic’, is now used to refer to the subject matter of the present book. Though the term

‘typology’ began its life with Gabelentz’s coinage at the turn of the twentieth century, linguistic typology itself has a much longer history, dating back to the eighteenth or even the seventeenth century. Needless to say, it is not possible to say exactly when linguistic typology began as a scholarly approach to the study of language. Like other scholarly approaches or scientific disciplines, linguistic typology underwent what Ramat (2011: 9) aptly calls ‘the incubation phase’, during which scholars pondered over problems or issues in general terms without realizing what they were thinking or writing about would eventually contribute to the development of an innovative way of investigating their objects of inquiry. One early example from this incubation phase of linguistic typology is the French abbot Gabriel Girard, who proposed a distinction between ‘analogous’ and ‘transpositive’ languages in the mid-eighteenth century: analogous languages have what we now call subject–verb–object order, while transpositive languages have different or even free word order. Moreover, analogous languages are claimed to mirror the ‘natural’ order of the thought process (the agent [= subject] exists, (s)he then does something [= verb], and, as a consequence, the patient [= object] is affected by the agent’s action). By contrast, transpositive languages do not reflect the ‘natural’ order of the thought process, displaying different or altered word orders, that is, word orders other than subject–verb–object. Moreover, analogous languages generally lack inflected forms, while transpositive languages are rich in inflected forms. Thus, a correlation can be drawn between these two typological properties: clausal word order and inflectional morphology. This correlation makes sense because in transpositive languages inflectional morphology encodes the distinction between subject and object (cf. head- and dependent-marking), thereby allowing the word order to deviate from the ‘natural’ order of the thought process. By contrast, the lack of inflectional morphology in analogous languages makes it important that the word order in these languages reflect the ‘natural’ order of the thought process. While it was more speculative than empirical, Girard’s study can perhaps be regarded one of the earliest instances of implicational typology since the presence of one property was utilized to draw inferences about the presence of another property. Put differently, Girard’s work was more than a classification of languages because of its attempt to explore a structural principle ‘that was capable of deeply characterizing languages from the point of view of’ the nature of human language (Ramat 1995: 45). For this reason, Ramat

(1995: 45) goes so far as to say that the French abbot and his immediate followers can be considered the true founders of linguistic typology, although Graffi (2011: 26) is of the view that Gabelentz should be regarded as the originator of linguistic typology, not least for his coinage of the term ‘typology’ and succinct description of linguistic typology. The reader is referred to Ramat (2011) for further examples of the incubation phase in the seventeenth and eighteenth centuries of linguistic typology.

Regardless of who was its founder, linguistic typology has been shaped over the centuries by many scholars, some of whom had lived before Gabelentz’s christening of the discipline. In common with other disciplines, the manner in which linguistic typology was conceptualized and developed was influenced by the intellectual milieu of particular historical periods in which it found itself. In the seventeenth century, the dominant intellectual movement was that of Rationalism (also known as the Enlightenment). The main driver of this intellectual movement was reason, instead of faith or tradition. Recall that the abbot Girard’s typology was based on the natural order of the thought process. This order, in turn, was claimed, through logical reasoning (or rationalism), to be the human mind’s way of thinking (or the natural flow of thought, as it were). Needless to say, this was highly speculative, since it was not backed by empirical evidence; rationalism alone would suffice for scholars of this intellectual background. Note that the focus of linguistic typology of this historical period was on the human mind’s natural (read: universal) way of thinking (i.e. unity), not so much on how different languages might or might not reflect the human mind’s way of thinking (i.e. diversity).

In the eighteenth and nineteenth centuries, the dominant intellectual milieu, in reaction largely to Rationalism, changed to Romanticism: emphasis was placed on human emotion or experience instead of reason. Not unexpectedly, language was also regarded as a human experience. For instance, speakers of different languages may have different experiences, which, in turn, may explain why they speak different languages in the first place. Scholars of this intellectual orientation went so far as to believe that language possessed an ‘inner form’. The inner form, in turn, was thought to be a manifestation of the spirit of the people (*Volksgeist*) who spoke the language (Greenberg 1974: ch. 3). In the words of Wilhelm von Humboldt (Finck 1899; Lehmann’s (1978c: 423) translation), ‘[t]he characteristic intellectual features and

the linguistic structure of a people stand in such intimacy of fusion with each other that if the one were presented the other would have to be completely derivable from it'. To wit, 'each language [was] a distinct revelation of the spirit (*Geist*)' (Greenberg 1974: 38). The inner form was also assumed to be reflected in 'variation in grammatical mechanisms employed in relating lexical concepts to each other [or relational meaning]' (Shibatani and Bynon 1995b: 4). This point of view led to the emergence of August von Schlegel's morphological typology, in which three basic strategies in the encoding of relational meaning were recognized: inflectional, agglutinative, and isolating—Wilhelm von Humboldt later added a fourth, incorporating, to Schlegel's tripartite classification.⁴ In inflectional languages, a single morpheme bears more than one meaning or represents more than one grammatical category. In agglutinative languages, a word may consist of more than one morpheme and may not be impervious to the conventional morpheme-by-morpheme analysis. In isolating languages, there is an equivalence relationship between word and morpheme with the effect that a morpheme can be taken to be a word or vice versa. In incorporating languages, the verb and its argument(s) may readily combine to create compound words. The unit of analysis in the morphological typology was undoubtedly the word, the structure of which 'was seized upon as in some sense central to the attempt to characterize the language as a whole' (Greenberg 1974: 36) so that 'the description of the entire grammatical system [could] be annexed to an exact description of the structure of the word in every language' (Lewy 1942: 15). Unlike the focus of Rationalism, that of Romanticism was laid squarely on linguistic diversity, and this interest in linguistic diversity was heightened by the discovery of 'exotic' languages outside the Old World through European imperial expansion (e.g. missionaries and traders). Unfortunately, the morphological typology came to be interpreted in highly subjective, evaluative terms with the effect that inflectional and isolating languages were regarded as the most and the least developed languages on an evolutionary scale, respectively. (This value-laden view relates to the development of so-called ethnopsychology

⁴ Comrie (1989: 45), on the other hand, adopts 'fusional' in lieu of 'inflectional' because 'both [agglutinative] and fusional languages, as opposed to isolating languages, have inflection, and it is... misleading to use a term based on (*in*)*flexion* to refer to one only of these two types'.

(*Völkerpsychologie*), which emphasized the relationship between language and thought.)

The nineteenth century witnessed the emergence of a profound intellectual paradigm, namely Darwinism. As is well known, this is a profound theory of biological evolution, proposing that all species of organisms emerge and develop through the natural selection of variations that individuals inherit, thereby enhancing their ability to compete, survive, and reproduce. This theory had an enormous impact on nineteenth-century scholars' view of language, as Bopp (1827: 1) avers:

Languages must be regarded as organic bodies [*organische Naturkörper*], formed in accordance with definite laws; bearing within themselves an internal principle of life, they develop and they gradually die out, after... they discard, mutilate or misuse... components or forms which were originally significant but which have gradually become relatively superficial appendages. (as translated by Sampson 1980: 17)

Under this Darwinian view, languages behave like biological species. (Try to read Bopp's remarks, with 'languages' replaced with the names of animals or plants.) Thus, just as biological species do, languages emerge, develop into different varieties, compete with other varieties or languages, and cease to exist. This apparent similarity was so remarkable that the linguist's language families, languages, dialects, and idiolects were thought to correspond to the biologist's genera, species, varieties, and individuals, respectively (Sampson 1980: 18). This evolutionary view of languages stood in stark contrast with the preceding Romanticist view, which was more humanities-based than scientific in that the latter regarded language as a subjective human experience, not as an entity to be described objectively as part of the natural world (Sampson 1980: 17). The Darwinian view of languages gave rise to a historical approach to the study of languages; not unexpectedly, the focus of this approach was based on the historical development or evolution of languages and the genealogical relationships among languages. In point of fact, this approach was strongly upheld at the time as the only natural one for the study of language and survived into the early twentieth century, as Meillet (1924) went so far as to claim that 'the only true and useful classification [of languages] is genetic [i.e. genealogical]' (Greenberg 1974: 40).

The historical approach was also augmented by mechanistic physics, another important scientific paradigm of the nineteenth century, albeit not as influential on the study of language as Darwinism. In

mechanistic physics, all natural phenomena could be understood by way of simple, deterministic laws of force and motion ‘so that all future states of the world could in principle be inferred from a complete knowledge of its present state’ (Sampson 1980: 17). (This is what Bopp had in mind when he mentioned ‘definite laws’ in his remarks cited above.) When extended to the study of languages, what this entails is that the history of language can also be described by simple, deterministic laws of sound changes. Scholars who fully embraced the two paradigms in their study of languages were so-called Neogrammarians (*Junggrammatiker*) in the latter half of the nineteenth century. They held the view that it was possible to ‘found a genuine science of laws [read: a scientific theory of language] based on rigorous methods and the discovery of sound law rested on historical comparison’ (Greenberg 1974: 39). Their primary objective was to discover sound laws that could account for the historical development of languages from their ancestors and their genealogical relationships. Thus, it comes as no surprise that unrelated languages fell outside the purview of Neogrammarians’ research by default, with the lack of genealogical relatedness leaving nothing to be taken care of by sound laws in the first place. Indeed, ‘[these scholars] saw no point to the comparison of unrelated languages’ (Greenberg 1974: 39); linguistic diversity, one of the primary pursuits of linguistic typology, was not on their research agenda. Needless to say, the exclusion of unrelated languages would considerably underrepresent the structural diversity of the world’s languages. Unfortunately, it was also assumed widely—albeit mistakenly—that typologically related languages were genealogically related as well. The dominant typological classification of this period was none other than the morphological typology mentioned earlier in relation to Romanticism; the morphological typology was, in fact, almost synonymous with linguistic typology during this period. So much so that anyone questioning or denying the connection between genealogical and typological relatedness was regarded as “heretical” (*Ketzerei*) at the time’ (Greenberg 1974: 59). Thus, linguistic typology was claimed to be of little or no use, because it could be subsumed, as it were, under the historical approach. Moreover, linguistic typology was not thought to be scientific enough as it could not reduce the nature of language to principles akin to Neogrammarian sound laws. Thus, linguistic typology came to be largely ignored in the latter half of the nineteenth century, falling by the wayside into near oblivion.

The first half of the twentieth century witnessed the rise in linguistics of structuralism (Saussure 1916), which brought about two major changes in the study of language. First, the focus shifted from diachrony, i.e. linguistic changes over time, to synchrony, i.e. the state of a language at a specific moment in time. Second, as the consequence of the first change, the historical approach, dominant in the latter half of the nineteenth century, found itself on the wane (for a useful discussion, see Sampson 1980: 21–33).⁵ These changes notwithstanding, the indifference to linguistic typology continued well into the first half of the twentieth century. Nonetheless, a small number of linguists in Europe as well as in the US maintained their interest in linguistic typology—in particular those associated with the Prague School in Europe.⁶ The European typologists were focused on what Greenberg (1974: 28–9) calls the generalizing goal of linguistic typology, that is, ‘the discovery of law-like generalizations in languages by drawing bounds to the concept “possible human language”’. Because of their anthropological—one may say, even Humboldtian—orientation, the few American typologists (especially Edward Sapir 1921: ch. 6), in contrast, were interested in discovering structural characteristics of individual languages or what Greenberg (1974: 28) refers to as the individualizing goal of linguistic typology. The reader may recall from the earlier discussion that the generalizing goal is somewhat akin to what seventeenth-century Rationalist philosophers of language had in mind (i.e. similarity), while the individualizing goal can be related to Romanticist scholars’ emphasis on individual languages as revelations of the spirit of their speakers (i.e. diversity). While the European typologists were well aware—at least far more so than the American typologists—of the connection between these two goals, especially in the form of implicational typology, it was not until the 1960s that serious attempts began to be made in linguistic typology in order to bring to the fore the linking of these two goals. The catalyst in this regard came when the Second World War forced some European linguists,

⁵ Greenberg (1974: 13) reports that the first occurrence of the word ‘typology’ in the linguistics literature was in the theses presented by the Prague linguists to the First Congress of Slavic Philologists held in 1928.

⁶ The majority of American structural linguists (e.g. Leonard Bloomfield) were not interested in linguistic typology because of its previous association with ethnopsychology (i.e. language in relation to thought), which, to their minds, was not suitable for empirical study. They were instead in favour of behaviouristic psychology, which deals with observable behaviour, not the unobservable in the mind.

including the Russian linguist Roman Jakobson (1896–1982), to flee or migrate to the US, where they (re)introduced linguistic typology, in particular implicational typology, to American linguists. However, linguistic typology went largely ignored, as it had been in Europe in the preceding century, until Joseph Greenberg (1963b) used the concept of implicational typology to investigate word order correlations, as briefly illustrated earlier in this chapter, in a relatively large number of languages, thereby elevating implicational typology in particular and linguistic typology in general to new heights. In other words, it was Greenberg who married up the two goals of linguistic typology more successfully than his predecessors and showed how to do implicational typology. In doing so, he ‘opened up a whole field of [linguistic–typological] research’ (Hawkins 1983: 23), revitalizing or resurrecting linguistic typology as a viable scientific approach to the study of language.

1.4 Concluding remarks

In this chapter, a brief description of how linguistic typology is carried out has been provided with special reference to what kinds of research question are raised, and how those questions are answered. Also provided is a historical overview of linguistic typology with special emphasis on how it has evolved conceptually over the centuries.

Study questions

1. Rapanui (an Oceanic language spoken on Easter Island) is a verb-initial language in that the verb appears before both the subject and the object, while Turkmen (a Turkic language spoken in Turkmenistan) is a verb-final language in that the verb follows both the subject and the object. In §1.2, the strong correlation between the verb position and the use of prepositions or postpositions was alluded to. According to that correlation, one may be able to predict that Rapanui and Turkmen use prepositions and postpositions, respectively. Check these predictions against the available data on Rapanui and Turkmen (e.g. Du Feu 1996; Clark 1998).
2. X and Y invested \$1,000 and \$3,000 in business, respectively. After a year in business, X and Y made \$3,000 and \$3,500 out of their investments,

respectively. Whose business do you think is more profitable, and why? Now, imagine you have two possible situations in your implicational typology research. In situation 1, you can state with confidence that if a language has properties A, B, and C, it also has property D. In situation 2, you can state with an equal amount of confidence that if a language has property E, it also has properties D, F, and G. Which implicational statement do you think is more valuable, and why? How would you characterize the common 'principle' underlying profitability in business and value in implicational typology?

3. Morphological typology, the most popular way of classifying languages in the period of Romanticism, was also used to evaluate the world's languages in terms of evolution. For instance, isolating languages were regarded as the most primitive or the least developed on the evolutionary scale. Even then, however, it was clear that Chinese flew right in the face of the evolutionary interpretation of the morphological typology. Why do you think this was so?
4. Ramat (1986: 8–9) points to diachronic dynamics of language: languages shift from Type X to Type Y while retaining some features of Type X. What kinds of problem do you think this poses for the Gabelentzian attempt to construct the entire language ('the entire lime tree') from one of its structural properties ('its leaf')?

Further reading

- Graffi, G. (2011). 'The Pioneers of Linguistic Typology: From Gabelentz to Greenberg', in J. J. Song (ed.), *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press, 25–42.
- Greenberg, J. H. (1974). *Language Typology: A Historical and Analytic Overview*. The Hague: Mouton.
- Horne, K. M. (1966). *Language Typology: 19th and 20th Century Views*. Washington, DC: Georgetown University Press.
- Ramat, P. (1995). 'Typological Comparison: Towards a Historical Perspective', in M. Shibatani and T. Bynon (eds.), *Approaches to Language Typology*. Oxford: Oxford University Press, 27–48.
- Ramat, P. (2011). 'The (Early) History of Linguistic Typology', in J. J. Song (ed.), *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press, 9–24.
- Song, J. J. (2010). 'Linguistic Typology', in K. Malmkjær (ed.), *Routledge Linguistics Encyclopedia*, 3rd edn. Abingdon: Routledge, 355–8.

Part I

Foundations: theory and method

2

Unity and diversity in the world's languages

2.1	Introduction	17
2.2	The connection between diversity and unity	20
2.3	The Principle of Uniformitarianism: a methodological frame of reference	22
2.4	When and where similarities count	26
2.5	Types of language universals and universal preferences	30
2.6	Concluding remarks	36

2.1 Introduction

References were made to Rationalism and Romanticism in Chapter 1, where the history of linguistic typology was discussed. These two schools of thought dominated the intellectual landscape in early modern Europe, the former in the seventeenth century and the latter in the eighteenth century and the first half of the nineteenth century. Not unexpectedly, they both had an enormous impact, in their respective historical periods, on how language was conceptualized as well as on how the study of language was approached and conducted. From these intellectual traditions emerged two overarching, albeit seemingly contradicting, research themes in the study of language: unity and diversity of human language. It was the unity of language that Rationalism highlighted, with particular reference to the human mind's natural (read: universal) way of thinking. Within Romanticism, in contrast, scholars drew attention to the diversity of human language, with language now

conceptualized as a manifestation of the spirit of its speakers—that is, different languages as reflections of different experiences. As will be shown in this chapter, these two themes continue to play an important role in modern linguistic typology, from Jakobson and Greenberg to the present day. While they are equally important concepts in linguistic typology (and indeed in any study of language for that matter), the pendulum has over the decades swung between the two themes. It was Jakobson and Greenberg who perceived the connection in linguistic typology between diversity and unity, while recognizing the search for the unity of human language as a loftier research goal than the diversity of human language. For instance, Jakobson (1958: 24) remarked: ‘Linguistic typology is an inference from the science of languages [read: diversity] for the science of language [read: unity].’ In other words, it is through the diversity of the world’s languages that one can arrive at the discovery of the unity of human language. Similarly, Greenberg (1974: 28–9) explains how ‘law-like generalizations in languages’ (read: unity) can be discovered on the basis of typological classifications (read: diversity) ‘by drawing bounds to the concept “possible human language”’. So much so that a typological classification is reduced to ‘an incidental by-product [that] is not of great interest for its own sake’ (Greenberg 1978a: 40). Put differently, diversity is what ‘provides material for establishing [unity]’ (Sanders 1976: 15; Mallinson and Blake 1981: 7). In the early stage of modern linguistic typology, therefore, the focus was on the discovery of language universals, which place the bounds on what is possible in human language. This ‘subjugation’, as it were, of diversity to unity in linguistic typology, however, has recently been rethought to the point of diversity attracting an increasingly greater amount of attention, especially in view of the realization that absolute language universals (i.e. all languages have X or there are no languages that lack X) are few and far between, and that it is, in point of fact, unrealistic to find them (e.g. Dryer 1992). So much so that linguistic typology should instead strive to discover what is probable, rather than what is possible, in human language. When the focus is placed on discovering what is possible vs impossible in language, the concept of language universal is based on the unity of human language (i.e. what is impossible in language will not be found in any known languages). In contrast, when the focus is shifted to what is probable in language, the concept of language universal must be attenuated to the effect that while X may be found in the majority of

the world's languages, there may also exist languages that possess Y or Z instead of X. In other words, attention is drawn to diversity, albeit with an eye to unity, since the majority of the world's languages may still fall into the type of X to the effect that one can still speak of a universal preference for X. This fundamental change in perspective, in turn, highlights the structural variation in the world's languages (e.g. not only X, but also Y and Z). More recently, the pendulum has swung even further in favour of diversity (e.g. Bickel 2007; Evans and Levinson 2009). In view of the unrealistic goal of discovering absolute language universals and the focus on 'what is probable in human language', it is now even more important to find out not only why the majority of the world's languages have X (i.e. universal preferences), but also why a small number of languages may have Y or Z in opposition to the universally preferred X—or, as Bickel (2007: 239) puts it, 'what's where why?'. This has motivated linguistic typologists to step outside the domain of language and explore historical, social, and cultural (i.e. non-linguistic) factors to account for the existence of languages that do not behave like the majority of the world's languages. For instance, languages with Verb–Object order (e.g. English, as in (1a), where *kissed* is V and *the man* is O) have a very strong tendency to place a Relative clause (enclosed in brackets) after the head Noun (underlined) that it modifies (e.g. English, as in (1b)), i.e. NRel order.

- (1) English (Germanic; Indo-European: UK)
 a. The woman kissed the man
 b. The woman [who kissed the man] is my sister.

Until recently, indeed, it was widely believed or assumed that VO languages always had NRel order. Then it was discovered (Dryer 1992, 2013e; but cf. Mallinson and Blake 1981: 285) that there are a small number of VO languages with RelN order in mainland China and Taiwan (i.e. Mandarin Chinese, Bai, and Amis), as exemplified in (2).

- (2) Chinese (Sinitic; Sino-Tibetan: China)
 a. tāmen tōu zìxíngchē
 3PL steal bicycle
 'They steal bicycles.'
 b. [wǒ gěi nǐ de] shū
 1SG give 2SG LINK book
 'the book that I gave you'

In other words, what was believed to be an absolute language universal (i.e. VO & NRel) has turned out to be slightly less than an absolute language universal. Note, however, that there is still a clear universal preference or a near language universal (Dryer 1992, 2013e): VO languages almost always have NRel order—98.8% of 421 VO languages sampled in Dryer (2013e) have NRel order indeed. Linguistic typologists have proposed that the uncommon combination of VO & RelN, as opposed to the global preponderance of VO & NRel, was due to Chinese having been influenced by OV languages with RelN order, located to the north of China (e.g. Mongolian and Tungus) and, subsequently, Chinese, now equipped with RelN order, influencing other VO languages that it came into contact with (Hashimoto 1986; Dryer 2003). Clearly, this structural similarity between Chinese and the neighbouring northern languages has its origins in historical, social, and/or cultural circumstances that brought them together and should thus be treated as contact-mediated.

2.2 The connection between diversity and unity

The area of linguistic typology dealing with diversity investigates the structural variation in the world's languages (i.e. everything that is attested), whereas the area of linguistic typology concerned with unity focuses on the discovery of language universals (i.e. what is possible) and universal preferences (i.e. what is probable). Language universals impose constraints or limits on structural variation within human language, while universal preferences delineate the dominance of some structural types over others. Typological investigation, in contrast, is concerned with classifying languages into different structural types. Thus, 'it may seem to the uninitiated something of a contradiction in terms to handle these apparently quite distinct areas of investigation together' (Mallinson and Blake 1981: 7; Greenberg 1986: 16). But, as may be gleaned from the preceding discussion, the contradiction is more apparent than real. The search for the unity in human language, in fact, builds crucially on the structural diversity in human language. This is because in order to discover language universals or universal preferences, what linguistic typologists first need is typological classification to work on. With languages classified into different types, linguistic typologists may be able to discern patterns or regularities

in the distribution of the types, for example, with some types being significantly more common than others, or with one (or more) of the logically possible types completely unattested or only marginally attested in the world's languages. Put simply, '[i]n order to understand Language, it is essential to understand languages' (Comrie 2006: 130).

This close relationship between language universals and universal preferences on the one hand, and linguistic typology on the other can be demonstrated by the preponderance of subject-initial order in the world's languages. According to Dryer (2013a), for instance, nearly 76.5% of his 1,377 sample languages have subject-initial word order, i.e. Subject–Object–Verb or Subject–Verb–Object. If the world's languages—or at least a significant portion of them—had not been surveyed in terms of all possible word orders (that is, not only subject-initial but also verb-initial and object-initial), this strong tendency would never have been brought to light in the first place. To put it differently, the typological classification of the world's languages in terms of the six word orders—(i) Subject–Object–Verb, (ii) Subject–Verb–Object, (iii) Verb–Subject–Object, (iv) Verb–Object–Subject, (v) Object–Verb–Subject, and (vi) Object–Subject–Verb—is a *sine qua non* for the significant generalization to be made about human language, that is, the majority of the world's languages have subject-initial word order. Imagine the prospect of discovering the universal preference in question by examining only one language or even a handful of languages! This may be too extreme an example but the point could not be made more strongly.

Further demonstration of the fruitful interaction between unity and diversity comes from another strong linguistic preference (albeit of an implicational nature): the presence of verb-initial order implying that of prepositions (see §1.2). This implicational statement also entails what is not possible in human language, namely the absence of verb-initial languages with postpositions (but see §2.3). Thus, the implicational statement does not only sanction verb-initial languages with prepositions as possible human languages but it also rules out verb-initial languages with postpositions as impossible human languages. Moreover, by making no negative claims about non-verb-initial languages, the implicational statement refers indirectly to the other two logical possibilities, namely non-verb-initial languages either with prepositions or with postpositions. In order to arrive at the actual formulation of this implicational statement, however, one must first ascertain

which of the four logical possibilities is attested or unattested in the world's languages. That can be achieved only on the basis of an initial typological classification of the languages of the world in terms of basic word order as well as the distribution of prepositions and postpositions. To wit, the search for the unity in human language is conducted on the basis of the structural diversity in human language. It is not possible to carry out the former without the latter.

The interaction between unity and diversity also highlights one of the virtues of formulating language universals or universal preferences on the basis of typological classification. Typological classification naturally calls for data from a wide range of languages (see Chapter 5 on how languages are selected or sampled for this purpose). Only by working with such a wide range of languages is one able to minimize the risk of elevating some of the least common structural properties to the status of language universals. This risk is more real than some linguists may be willing to admit because, when deciding to work with a small number of familiar or well-known languages (for whatever reasons), one is likely to deal with structural properties which may not in any real sense be representative of the world's languages.¹ For instance, use of relative pronouns is very common in European languages but it has turned out to be a cross-linguistically infrequent phenomenon (Comrie 1989: 149; Comrie 2006; Comrie and Kuteva 2013c). Therefore, universal claims about, or universal theories of, relative clauses which are put forth on the basis of these European languages alone should immediately be suspect.

2.3 The Principle of Uniformitarianism: a methodological frame of reference

Strictly speaking, to claim that something is a language universal or a universal preference is patently premature, if not meretricious, when one thinks about the current level of documentation among the world's languages. There are reported to be nearly 7,105 languages in the world

¹ For instance, Chomsky (1980: 48) avers: 'I have not hesitated to propose a general principle of linguistic structure on the basis of observation of a single language... Assuming that the genetically determined language faculty is a common human possession, we may conclude that a principle of language is universal if we are led to postulate it as a "precondition" for the acquisition of a single language.'

(Ethnologue, 2013), but unfortunately, ‘[l]ess than 10% of these languages have decent descriptions (full grammars and dictionaries)’ (Evans and Levinson 2009: 432). (Incidentally, this highlights the urgent need to document languages before they disappear into oblivion (e.g. Nettle and Romaine 2000; Grenoble and Whaley 2006; Evans 2010).) What this means is that language universals or universal preferences proposed so far are all based on less than 10% of the world’s languages, even if we assume, for the sake of argument, that all documented languages have been taken into consideration for every language universal or universal preference that has ever been proposed. Over 90% of the world’s languages, which await (proper) documentation, remain to be brought into the fold of linguistic research. Put simply, proposed language universals or universal preferences should never be understood to be established facts about languages but should instead be taken to be nothing more than observations or hypotheses based on what has been documented about the world’s languages. In other words, proposed language universals or universal preferences are in need of verification against further data as they become available. A case in point is one of the language universals proposed by Greenberg (1963b): ‘Languages with dominant VSO order are *always* prepositional’ (emphasis added). Subsequent research (e.g. Dryer 1991: 448), however, has revealed that this is not entirely correct. There are languages, albeit in small numbers, that have both VSO order and postpositions, e.g. Cora (Corachol; Uto-Aztecan: Mexico), Ewe (Kwa; Niger-Congo: Togo and Ghana), Finnish (Finnic; Uralic: Finland), Guajajara (Tupi-Guaraní; Tupian: Brazil), Kashmiri (Indic; Indo-European: Pakistan and India), Northern Tepehuan (Tepiman; Uto-Aztecan: Mexico), and Waray (Warayic; Gunwinyguan: Australia). When Greenberg proposed this universal—and his other universals for that matter—in the early 1960s, he never intended them to be facts about the world’s languages, but rather generalizations based on the data available to him at the time. Put differently, Greenberg’s language universals were merely hypotheses or ‘summar[ies] of data observed to date’ (Dryer 1998: 143). In this respect—this is an important thing to remember—linguistic investigation is no different from any other scientific investigation. Every scientific statement, whether in physics or genetics, is simply a summary of data observed to date, or a generalization or a hypothesis based on that summary. More to the point, no scientific investigation can examine every instance of what it

is investigating. Typically, a small sample of the population under investigation is drawn for purposes of formulating a generalization or a hypothesis, which will continue to be tested against further data as they are collected (see §5.6 for sampling in linguistic typology). This important point should not be lost sight of when language universals or universal preferences are spoken of. To wit, the validity of language universals and linguistic preferences can only be strengthened or weakened by means of further empirical testing.

In order to arrive at a proper understanding of the nature of human language, it is not sufficient to study only living languages—even if it is possible to include all living languages in one's investigation. Ideally, linguistic typologists must study extinct languages as well. Indeed, linguistic typologists often study not only currently spoken languages but also extinct languages. This may perhaps strike one as odd, if not surprising, because one may expect typological classification to be concerned only with the currently spoken languages of the world. One may be inclined to think that language universals represent constraints or limits on structural variation within human language as it *is*, not as it *was* (or for that matter as it *will be*). Why do linguistic typologists also include extinct languages in their investigation? The assumption underlying this inclusion is what is generally known as the *Principle of Uniformitarianism* in linguistics (see Lass (1980: 53–7, 1997: 24–32) for discussion thereof in the context of historical linguistics).² Basically, what it means is that human languages of the past—or of the future for that matter—are not essentially different in qualitative terms from those of the present. This principle claims, therefore, that the fundamental properties of human language have remained invariable over time. There are believed to be no differences in evolutionary terms between languages of the past—as far back as one can go and claim the existence of human languages—and those spoken today. In other words, human language of today is at the same level of evolution as that of, say, 60,000 years ago. Thus, in order to come to grips with the full structural diversity of human language, researchers must also investigate not only living languages but also extinct ones. Imagine a possible situation

² This principle was first introduced into the study of language by Neogrammarians from the natural science thesis of Hutton and Lyell. Karl Brugmann is quoted as saying (Collinge 1995: 1561): '[t]he psychological and physiological nature of man as speaker must have been essentially identical at all epochs.'

in which particular structural types unattested in living languages happen to have existed only in languages that are no longer spoken.

The Principle of Uniformitarianism is, of course, something that has never been subjected to empirical verification and cannot be put to the test for obvious reasons; one simply cannot go back in time and examine languages spoken, say, 60,000 years ago to see whether or not they were qualitatively the same as those of today. Nor is there any logical reason why the principle should be correct. Nonetheless it plays an important role in linguistic typology (and equally in historical linguistics). The primary aim of linguistic typology is to discover universal properties or preferences of human language. If human languages were spoken 60,000 years ago, then, these languages must also be included in any typological study, which is utterly impossible. To get an idea of the linguistic diversity in the past, one can refer to Evans and Levinson (2009: 432), who suggest that 500 years ago, before European colonization began, there were probably twice as many languages as there are now, and to Pagel (2000: 395), who claims that over half a million languages have ever been spoken on this planet, if humans began talking 200,000 years ago and languages evolved at a rate of one per 500 years. In the absence of the Principle of Uniformitarianism, then, no typological analysis will be possible or, more accurately, complete simply because it is impossible to 'recover' all unrecorded, extinct languages from oblivion. With the Principle of Uniformitarianism in place, however, linguistic typologists can examine languages spoken today and, if and where possible, extinct but documented languages as well and can still make statements or generalizations about the nature of human language. Similar comments can also be made of languages of the future. Since it is expected that they will also be human languages, any typological study must in principle include them as well, which is out of the question. But the Principle of Uniformitarianism also works in the opposite direction of time from the present, thereby allowing linguistic typologists to extend to languages of the future what universal properties or preferences that they may have discovered on the basis of currently available data. While the complete structural diversity of human language will be impossible to capture, what is true of today's human language can be assumed to be true of yesterday's and tomorrow's human language. After all, under the Principle of Uniformitarianism the nature of human language is assumed not to change qualitatively over time.

There are also practical reasons why the Principle of Uniformitarianism is adhered to in linguistic typology. Without this principle, languages must be seen to evolve constantly as time passes by. But if languages were evolving through time, and were conceived of as being at different stages of linguistic evolution, grammatical descriptions that linguistic typologists employ for their research would be completely useless for typological research because they invariably—and inevitably—record languages at different points in time or at different stages of evolution, with some grammars being descriptions of languages of more than a few hundred years ago, and others being far more recent ones.

The absence of the Principle of Uniformitarianism will also lead to the view—which incidentally is generally not accepted in linguistics—that some languages should be at a more advanced stage of evolution than others because one would not be able to claim that all human languages have evolved to the same level (see §1.2 on an evolutionary interpretation of the morphological typology in the nineteenth century). If languages were at different stages of linguistic evolution, it would be impossible to engage in any typological research since one would (arbitrarily) have to target at one particular stage of evolution that all human languages have reached at one time or another, and to study all grammatical descriptions of the world's languages at *that* stage of evolution (assuming, of course, that it is possible to select such a stage, and also to have access to all grammatical descriptions at once).³

To sum up, the Principle of Uniformitarianism provides a kind of frame of reference within which typological research can be carried out productively without being hindered unduly by the intractable methodological issue, which does not necessarily have to be resolved—and most probably never will—at the current stage of development of linguistic typology as an empirical approach to the study of language.⁴

2.4 When and where similarities count

When studying the structural diversity of the world's languages with a view to uncovering the unity of human language, linguistic typologists

³ A related question will be: which stage of evolution in human language should be chosen as the 'target' stage?

⁴ One may choose to use the descriptive label 'an escape hatch', rather than 'a frame of reference'.

must take care to separate language universals or universal preferences from structural similarities brought about by non-linguistic factors, e.g. historical accidents. Imagine a hypothetical world where there are 1,000 languages (loosely based on Dryer 1989: 259–60). In this world, there is one large language family of 900 languages, with the remaining 100 languages evenly distributed among ten small language families (i.e. ten languages in each of the small language families). All the 900 languages in the large family have type X, and the languages of the other ten small families all have type Y. Now, is it safe to conclude from this distribution that there is a universal preference for X over Y, since X is attested in 900 languages (90%) while Y is attested in only 100 languages (10%)? This answer is no, because of the fact that X is found in only one language family and Y in the remaining ten families: Y is attested in far more language families than X is. The fact that it could have been the other way round (i.e. X in ten small language families and Y in one large family) suggests that the distribution of X and Y in the hypothetical world's languages may well be a historical accident, having nothing to do with what is or is not linguistically preferred. For further illustration, imagine that in the distant past, the large language family used to have ten languages, with the remaining 990 languages evenly distributed among the other ten language families (i.e. 99 languages per family). Through their superior technology, however, the size of the former language family subsequently increased exponentially at the expense of the latter language families. (Technologically advanced people have better access to natural resources, increasing not only their chances of survival but also the size of their territory; for a recent example, one can point to the spread of Europeans and their languages, e.g. English and Spanish, to other parts of the world.) Thus, the presence of X in the majority of the hypothetical world's languages (i.e. the large language family) is caused by the technological superiority of their speakers, and certainly not by the universal linguistic preference for X over Y. From this, it is glaringly obvious that language universals or universal preferences cannot be established on the basis of structural similarities brought about by such historical accidents. This is why it is decided that X is not the universal preference in spite of the fact that it is attested in 90% of the hypothetical world's languages. The reader may wonder, at this point, whether the preponderance of subject-initial word order in the real world's languages, alluded to earlier as a universal preference, may also be the outcome of a similar historical accident,

namely the economic, social, and political domination of speakers of subject-initial languages over those of non-subject-initial languages. This is a valid point. Indeed, linguistic typologists have devised methods of ascertaining whether the global dominance of subject-initial word order (or other structural properties for that matter) is a genuine universal preference or the outcome of a historical accident (for detailed discussion, see §5.7).

There are three major ways languages come to have similar properties: (a) shared genealogical origin, (b) language contact, and (c) language universals or universal preferences. Linguistic typology is concerned primarily with (c), while not neglecting to pay attention to (a) and (b), especially when also explaining 'exceptions' to language universals or universal preferences. Thus, when and where unity is the focus of investigation, it is (c) that counts, and (a) and (b) have to be taken out of consideration. Needless to say, however, if the focus of investigation is 'what's where why' à la Bickel (2007: 239), (a) and (b) should also be brought into the picture.

The make-believe example given above illustrates shared genealogical origin. The fact that there are 900 languages in the large language family with property X is due to the fact that these languages all derive historically from one ancestral language. In other words, languages may have similar structural properties because they have inherited them from their common ancestors. Languages may not begin their lives as independent languages but as dialects of single languages, e.g. French, Italian, Portuguese, Romanian, and Spanish, derived from Latin. In linguistics, this genealogical relationship is typically captured by analogy with a biological relationship, that is, parent and daughter languages. For instance, Latin is the parent language of French, Italian, Portuguese, Romanian, and Spanish, or conversely, French, Italian, Portuguese, Romanian and Spanish are the daughter languages of Latin. Further back in time, Latin was part of yet another ancestral language, and so on. Using what is known as the Comparative Method (e.g. Campbell 2013), historical linguists have classified the world's languages into a number of language families (although they may disagree about some of them). For instance, English, French, Greek, Irish, Russian, Hindi, and Urdu all belong to one and the same language family called Indo-European; these modern languages are descendants of a single language spoken some 6,000 years ago, with its reconstructed name, Proto-Indo-European.

The case of shared properties through language contact has already been alluded to when Chinese and its neighbouring languages were discussed earlier with respect to the universal preference for VO and NRel order. Chinese belongs to the Sino-Tibetan family, whereas the neighbouring northern languages belong to the Mongolic or Tungusic branch of the Altaic family. However, Chinese and these northern languages share structural properties including RelN order, in spite of the difference in their basic word order (for detailed discussion, see Dryer 2003: 48–53). In the case of Chinese, a VO language, its RelN order goes against the grain of the universal preference, i.e. VO & NRel, as it were. It has already been explained that Chinese adopted RelN order from the northern OV languages in preference to NRel order. This illustrates how languages of different genealogical origins may end up with common structural properties through contact, in opposition to language universals or universal preferences. Another example of this kind comes from the Asia Minor Greek dialects, spoken in the regions of Sílli, Cappadocia, and Phárasa. These Asia Minor Greek dialects were heavily influenced by Turkish through prolonged contact. Greek is an Indo-European language, while Turkish is a non-Indo-European language or a Turkic language. One of the Turkish structural features imported into the Greek dialects in question is RelN order. In Greek (3), a VO language, the relative clause follows the head noun (i.e. NRel order), as expected of VO languages (i.e. the universal preference for VO & NRel), but in Sílli dialects (4), for instance, the converse is frequently the case, that is RelN, just as in Turkish (5), an OV language.

- (3) Greek (Hellenic; Indo-European: Greece)

tò peðì pû to ìða
 the boy COMP it saw-I
 ‘the boy who(m) I saw’

- (4) Sílli Greek (Hellenic; Indo-European: Turkey)

kját íra perí
 COMP saw-I boy
 ‘the boy who(m) I saw’

- (5) Turkish (Turkic; Altaic: Turkey)

gör-düğ-üm oğlan
 see-NOMN-1.SG.POSS boy (son)
 ‘the boy who(m) I saw’

Such contact-mediated structural similarities should not come into the picture when language universals or universal preferences are identified or delineated. If the political situation in the region in the fifteenth to early twentieth century had been different, i.e. Greek speakers' domination over Turkish speakers instead of the converse, it could easily have been Turkish 'borrowing' NRel order from Greek. Languages may come to share structural properties through contact because speakers of languages may adopt them from other languages that they come into contact with, not necessarily because those properties are inherently preferred in human languages. Even if borrowed properties are universal preferences, the structural borrowing is brought about first and foremost through contact between languages. To wit, contact-mediated similarities also are due to historical accident, just as similarities through shared genealogical origins are.

This leaves the third major way languages come to share structural properties: language universals or universal preferences. The world's languages or at least the majority of the world's languages may have such and such structural properties because they are due to the very nature of human language: all other things being equal, languages must have such and such properties because they are what makes human language what it is. For instance, there is a clear, strong tendency for VO languages to have NRel order. This correlation is attested in virtually all VO languages of the world (with a very small number of exceptions, which can be explained by reference to language contact, as has already been noted). This near-universal correlation must then have to do with the nature of human language: if a language opts for VO order, it must also opt for NRel order. The discovery of the nature of human language is the primary goal of linguistic typology, and in fact, any linguistic theory for that matter. The primary research goal is to uncover the nature of human language, with an eye to accounting for cases where the nature of human language may be 'suspended' or 'put on hold' because of overriding historical, social, or cultural circumstances.

2.5 Types of language universals and universal preferences

Properties such as the preponderance of subject-initial languages in the world's languages are often referred to as language universals in the

literature. Strictly speaking, however, language universals must be true of all languages. Under this strict definition of the term, the property of subject-initial order does not qualify as a language universal since it is only a tendency in human language, albeit a very strong one. In other words, only properties which all human languages have in common may be taken to be language universals. This is why the tendency for subject-initial order has been referred to in the present book as a universal preference, not as a language universal: a linguistic preference attested very commonly or widely in the world's languages.⁵ The correlation between verb-initial word order and prepositions is another universal preference in that there are only a small number of languages with verb-initial order and postpositions, with the majority of verb-initial languages being prepositional. In the literature, language universals, as interpreted strictly as exceptionless, are referred to as absolute language universals, whereas universal preferences, admitting of a small number of exceptions, are called non-absolute or statistical language universals. (Bear in mind that in this commonly used distinction, language universals are liberally interpreted as including not only absolute language universals but also universal preferences.)

Absolute language universals are very hard to find. It is not the case that they do not exist. They certainly do. But they are not numerous and they tend to be 'banal' or 'trite' (Greenberg 1986: 14, 15). For instance, the fact that all languages have vowels has been proposed as an absolute language universal. This, however, is rather uninspiring. One may ask: all language have vowels, so what? It does not seem to lead to any further interesting questions about human language, except for the question as to why all languages must have vowels. More seriously, Evans and Levinson (2009: 438) correctly point out that even this seemingly absolute language universal is not without exceptions (in this case, many), when sign languages are taken into account! Experience shows that what was proposed as a (possible) absolute language universal almost always turns out to have exceptions. The erstwhile absolute language universal involving VO and NRel order has already been mentioned as a similar case in point. Yet another example of

⁵ Nichols (1992: 42) describes universal tendencies as properties or correlations favoured in languages independent of geography and genetic affiliation, and thus as universal preferences in the world's languages.

this kind comes from Greenberg (1966: 50): all languages mark the negative by adding some morpheme to a sentence. However, Evans and Levinson (2009: 439) point to Classical Tamil (Southern Dravidian; Dravidian: India and Sri Lanka) as a counterexample, as this language marks the negative by deleting the tense morpheme present in the positive. One may argue that Classical Tamil is only one counterexample among the world's 7,000-odd languages and brush it aside. (At the same time, this shows that one exception is all it takes to turn an absolute language universal into a non-absolute statement.) However, one must realize that only less than 10% of the world's languages have adequate descriptions. One cannot be sure whether there may be other yet-to-be-documented (or even yet-to-be-discovered) languages that behave exactly like Classical Tamil. More frequently than not, absolute language universals have been formulated on the basis of an even smaller number of languages. As more and more languages become documented and brought to the attention of researchers, new properties or strategies are very likely to show up, flying in the face of absolute language universals or universal preferences. Thus, it does not come as a surprise that absolute language universals are hard to come by, and virtually all absolute language universals claimed so far have turned out to be less than what they were initially thought to be. Moreover, one must bear in mind, as Evans and Levinson (2009: 439) do, that 'the relevant test set is not the 7,000 odd languages we happen to have now, but the half million or so that have existed, not to mention those yet to come [into existence]'. In view of this reality, not surprisingly, the focus of linguistic typology has recently been shifted from unity to diversity, as the clear message coming from the world's languages, as more and more of them become documented, is that there is always going to be a language somewhere that will throw a typological curve ball, as it were. Thus, some linguistic typologists (e.g. Bickel 2007; Evans and Levinson 2009) argue that more effort should instead go into documenting the structural diversity of the world's languages before advancing premature claims about the nature of human language. But at the same time, it must be borne in mind that making observations or hypotheses about human language on the basis of available data is legitimate business, not just in linguistic typology but also in other scientific disciplines.

The fact that there are hardly any absolute language universals that can stand the test of time does not detract from the fact that there is a

substantial degree of unity in the world's languages. Various universal preferences capture the very unity of human language. A small number of languages that deviate from this near unity may further reflect the structural diversity in the world's languages, and must be accounted for in whatever way they can. As has already been pointed out on more than one occasion, however, these 'deviations' tend to have social, cultural, and/or historical reasons behind them. These non-linguistic reasons enable linguistic typologists to understand why property X exists in language Y at a particular point in time, in opposition to the overwhelming global tendency. In other words, it is important to ask why and how the 'deviating' languages have arisen. Moreover, it is important to answer these questions because in doing so, linguistic typologists will find themselves in a stronger position to strengthen the validity of universal preferences. When exceptions to universal preferences are addressed on their own terms, the conceptual as well as empirical strength of proposed universal preferences is by no means vitiated but rather increased to a greater extent than would otherwise be the case. This is because exceptions have valid reasons for being exceptions. Put differently, a small number of exceptions to universal preferences are not really counterexamples as such, but rather the outcome of non-linguistic factors 'interfering' with linguistic preferences (read: the unity of human language). In this respect, exceptions to universal preferences represent non-linguistic variables that may override linguistic preferences in highly specific historical, social, and cultural contexts.

As has been shown, linguistic statements about the nature of human language can be formulated by using two parameters: (a) absolute vs non-absolute; and (b) implicational vs non-implicational. Absolute statements are exceptionless by definition. An example of this type of universal is: all languages have ways to turn affirmative sentences into negative ones (e.g. *James kicked the dog* → *James did not kick the dog*). Non-absolute statements—also known as statistical universals or, as in this book, universal preferences—are not without exceptions but the empirical strength of this type of statement far outweighs the number of exceptions that may exist. The stated preponderance of subject-initial order in the world's languages is a non-absolute statement. Various statistical methods are employed in order to determine whether or not a given tendency is statistically significant (see §5.6).

Implicational statements take the form of 'if *p*, then *q*'. The presence of one property (i.e. the *implicans*) implies that of another (i.e. the