Edited by Timothy Shopen

Language Typology and Syntactic Description

Volume III: Grammatical Categories and the Lexicon

SECOND EDITION

CAMBRIDGE

CAMBRIDGE www.cambridge.org/9780521581585

This page intentionally left blank

Language Typology and Syntactic Description

Second edition

Volume III: Grammatical Categories and the Lexicon

This unique three-volume survey brings together a team of leading scholars to explore the syntactic and morphological structures of the world's languages. Clearly organized and broad-ranging, it covers topics such as parts of speech, passives, complementation, relative clauses, adverbial clauses, inflectional morphology, tense, aspect mood, and deixis. The contributors look at the major ways that these notions are realized, and provide informative sketches of them at work in a range of languages. Each volume is accessibly written and clearly explains each new concept introduced. Although the volumes can be read independently, together they provide an indispensable reference work for all linguists and field workers interested in cross-linguistic generalizations. Most of the chapters in the second edition are substantially revised or completely new – some on topics not covered by the first edition. Volume III covers typological distinctions in word formation; lexical typologies; inflectional morphology; gender and noun classes; aspect, tense, mood; and lexical nominalization.

Timothy Shopen (1936–2005) was Senior Lecturer in Linguistics at the Australian National University. He had over forty years' experience of teaching and researching a variety of the world's languages, and also held posts at Indiana University and the Center for Applied Linguistics in Arlington, Virginia. In addition to *Language Typology*, he was editor of *Standards and Dialects in English* (1980), *Standards and Variables in English* (1981), *Languages and their Speakers* (1987), and *Languages and their Status* (1987).

Language Typology and Syntactic Description

Second edition
Volume III: Grammatical Categories and the Lexicon

Edited by

Timothy Shopen[†]

CAMBRIDGE
UNIVERSITY PRESS

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

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK
Published in the United States of America by Cambridge University Press, New York
www.cambridge.org
Information on this title: www.cambridge.org/9780521581585

First published in print format 2006

ISBN-13 978-0-511-35306-2 eBook (Adobe Reader)
ISBN-10 0-511-35306-5 eBook (Adobe Reader)

ISBN-13 978-0-521-58158-5 hardback
ISBN-10 0-521-58158-3 hardback

ISBN-13 978-0-521-58855-3 paperback
ISBN-10 0-521-58855-3 paperback

Contents

List of figures *page* xi
List of tables xii
List of contributors xiv
Acknowledgements xv
List of abbreviations and symbols xvii

1 Typological distinctions in word-formation 1
ALEXANDRA Y. AIKHENVALD

0 Introduction 1
1 The word 1
2 Morphological typology and word-formation 3
2.1 Transparency of word-internal boundaries 3
2.2 Internal complexity of grammatical words 5
2.3 Integrating the two parameters 8
2.4 Word-formation and syntax in languages of different types 9
3 Noun incorporation 11
3.1 Formal properties of incorporation 12
3.1.1 What material gets incorporated 12
3.1.1.1 (I) Incorporation of a free form of a noun 12
3.1.1.2 (II) Incorporation of a bare noun root 12
3.1.1.3 (III) Incorporation of a suppletive or reduced stem 13
3.1.1.4 (IV) Incorporation of the whole NP 13
3.1.2 The degree of formal cohesion between components 14
3.2 Functional types of incorporation 15
3.2.1 Type 1. Lexical compounding 15
3.2.2 Type 2. The manipulation of case 16
3.2.3 Type 3. The regulation of information flow 17
3.2.4 Type 4. Incorporation of modifiers 17
3.2.5 Type 5. Classificatory incorporation 17
3.3 Syntactic functions of incorporated nouns, and their incorporability 19
4 Structure and iconicity in word-formation 21
5 Compounding 24
5.1 How to distinguish compounds from phrases 24
5.1.1 (I) Phonological criteria 25
5.1.2 (II) Morphological criteria 26

5.1.3 (III) Morphosyntactic criteria 27
5.1.4 (IV) Semantic criteria 28
5.2 Nominal compounds 28
5.2.1 (I) Endocentric, exocentric and coordinate compounds 30
5.2.2 (II) Root compounds and synthetic compounds 31
5.3 Verbal compounds 32
5.4 Compounding in other word classes 34
6 Derivation 35
6.1 Inflection and derivation 35
6.2 Roots, stems and affixes 38
6.3 Types of derivational processes 40
6.3.1 Functional classification of derivational devices 40
6.3.2 Formal classification of derivational devices 44
7 Productivity and related phenomena in word-formation 49
7.1 Determining productivity 50
7.2 Factors conditioning productivity 50
7.3 Factors restricting productivity 51
7.3.1 (I) Phonological factors 52
7.3.2 (II) Morphological and morphosyntactic factors 52
7.3.3 (III) Semantic and pragmatic factors 53
7.3.4 (IV) Lexical factors 53
7.4 Lexicalization and predictability 54
7.5 Loss and gain of productivity 56
7.6 Productivity and creativity: hierarchy of productivity 57
8 Grammaticalization and lexicalization in word-formation 58
8.1 Grammaticalization in word-formation 58
8.2 Lexicalization in word-formation 60
9 Conclusions 61
10 Suggestions for field workers in describing types of word-formation 62
10.1 Questions to ask 63
11 Suggestions for further reading 64

2 Lexical typologies 66
LEONARD TALMY

0 Introduction 66
0.1 Characteristics of lexicalization 68
0.2 Sketch of a motion event 70
1 The verb 72
1.1 Motion + Co-event 72
1.1.1 The pattern underlying Co-event conflation 74
1.1.2 Properties of Co-event conflation 75
1.1.2.1 Two verb usages 75
1.1.2.2 The lexicalization account 76
1.1.2.3 Translational and self-contained Motion 79
1.1.3 Extensions of the Co-event conflation pattern 81
1.1.3.1 Conflation onto mid-level verbs based on BE_{LOC} or MOVE 82
1.1.3.2 Conflation onto combinations of MOVE with matrix verbs 83

1.1.3.3 Conflation onto metaphorically extended MOVE 84
1.1.3.4 Conflation across the various relations of the Co-event to the Motion event 85
1.1.3.5 Multiple conflation 87
1.2 Motion + Path 88
1.3 Motion + Figure 96
1.4 A typology for motion verbs 99
1.4.1 Motion + Co-event, Path, or Figure 99
1.4.2 Motion + Ground 99
1.4.3 Motion + two semantic components 100
1.4.4 Motion + no further semantic component 101
1.4.5 Motion + a minimally differentiated semantic component 102
1.4.6 Split system of conflation 103
1.4.7 Parallel system of conflation 104
1.4.8 Intermixed system of conflation 105
1.5 Aspect 107
1.6 Causation 108
1.7 Interaction of aspect and causation 117
1.7.1 Consistency of patterns within a language 123
1.7.2 Other aspect–causative types 126
1.8 Personation 128
1.9 Valence 131
1.9.1 General considerations 131
1.9.2 Valence in verbs of affect 134
2 Satellites 138
2.1 Path 141
2.2 Path + Ground 146
2.3 Patient: (Figure/)Ground 149
2.4 Manner 150
2.5 Cause 151
2.6 Motion-related satellites extending the motion typology 153
2.6.1 Verb-framed and satellite-framed systems 153
2.6.2 Typological shift and maintenance 154
2.7 Aspect 155
2.8 Valence 158
2.8.1 Satellites determining the Figure–Ground precedence pattern of the verb 158
2.8.2 Satellites requiring Direct Object to indicate 'bounded Path' 161
3 Salience in the verb complex 163
4 Conclusion 167
5 Suggestion for further reading 168

3 Inflectional morphology 169
BALTHASAR BICKEL AND JOHANNA NICHOLS
0 Introduction 169
1 Formatives and morphological types 172
1.1 Words versus formatives 172
1.2 Clitics 174

1.3 Degree of fusion 180
1.3.1 Isolating 180
1.3.2 Concatenative (bound) 181
1.3.3 Nonlinear 182
1.4 Flexivity (variance, lexical allomorphy, inflectional classes) 184
1.5 Semantic density 188
1.5.1 Exponence 188
1.5.2 Synthesis and wordhood 189
2 Locus 193
3 Position 197
4 Paradigms 201
4.1 Inflectional classes 202
4.2 Syncretism 207
4.3 Defectivity and suppletion 208
4.4 Deponence 208
4.5 Eidemic resonance 209
4.6 Case inventories and case terminology 210
5 Markedness and obligatoriness 212
6 Layered (hierarchical) versus templatic morphology 214
7 Two examples of common inflectional categories: person and number 220
7.1 Person 220
7.1.1 Exclusive versus inclusive 220
7.1.2 Conjunct/disjunct systems 223
7.1.3 Person and the indexability hierarchy 224
7.2 Number 227
8 Morphology in syntax 229
8.1 Agreement 229
8.2 Case spreading and stacking 235
9 Conclusions 239
10 Suggestions for further reading 239

4 Gender and noun classes 241
GREVILLE G. CORBETT
0 Introduction 241
1 Terms and analysis 242
1.1 Analysis based on agreement classes 243
1.2 Classifiers and complex systems 253
2 The speaker's problem: gender assignment 258
2.1 Semantic assignment 259
2.2 Predominantly semantic assignment 259
2.3 Morphological assignment 261
2.4 Phonological assignment 264
3 Default genders 266
3.1 Types of default 267
3.2 Defaults in gender systems 268
4 Gender resolution 273
4.1 Semantic gender resolution 273
4.2 Syntactic gender resolution 274
4.3 Mixed semantic and syntactic gender resolution 275

4.4 The relation between resolution and assignment 276
5 Prospects 276
6 Suggestions for further reading 279

5 Aspect, tense, mood 280
ALAN TIMBERLAKE
0 Introduction 280
1 Aspect 284
2 Tense 304
3 Mood and modality 315
4 Aspect, tense, and modality, in text and in general 330
5 Suggestions for further reading 332

6 Lexical nominalization 334
BERNARD COMRIE AND SANDRA A. THOMPSON
0 Introduction 334
1 Processes for forming nouns from lexical verbs and adjectives 335
1.1 Action/state nominalization 335
1.2 Agentive nominalization 336
1.3 Instrumental nominalization 338
1.4 Manner nominalization 339
1.5 Locative nominalization 340
1.6 Objective nominalization 340
1.7 Reason nominalization 342
1.8 Predictability and productivity 342
2 Processes for forming Noun Phrases from predicates and propositions 343
2.1 The 'action nominal' 343
2.1.1 Verbal and nominal categories 345
2.1.1.1 Verbal categories 345
2.1.1.1.1 Tense 346
2.1.1.1.2 Aspect 347
2.1.1.1.3 Voice 348
2.1.1.1.4 Transitivity 351
2.1.1.1.5 Negation 352
2.1.1.1.6 Summary 353
2.1.1.2 Nominal categories 353
2.1.2 Syntactic collocation 355
2.1.2.1 Valency 355
2.1.2.1.1 Subjects and objects assimilate to NP syntax 355
2.1.2.1.2 Subjects and objects retain sentence syntax: Tamil and Avar 362
2.1.2.1.3 Subjects and objects only partially assimilate to NP syntax 363
2.1.2.1.4 Unexpressed subjects 368
2.1.2.1.5 Idiosyncrasies in valency of action nominals 369
2.1.2.2 Adverbs and adjectives 374

2.2 Nominalizations with no lexically derived noun 376
2.3 Functions of nominalizations 377
3 Devices for forming nouns from nouns 379
3.1 Abstract nouns 379
3.2 Concrete nouns: augmentative/pejorative/diminutive 380
4 Summary 381
5 Suggestions for further reading 381

Bibliography 382
Language index 411
Subject index 416

Figures

1.1 Interaction of two types of parameters in word-formation *page* 8
2.1 Co-event conflated in the Motion verb 72
2.2 Path conflated in the Motion verb 89
2.3 Figure conflated in the Motion verb 96
2.4 Aspectual meanings lexicalized in verb roots 106
4.1 The gender system of Romanian 247
4.2 The gender system of French 248
4.i The Agreement Hierarchy 252
5.1 Pieter Bruegel, *The Hunters in the Snow* (1565) 281

Tables

1.1 Morpheme slots in Tiwi verb (Lee (1987:152–5)) *page* 7
1.2 Reclassification of an inanimate noun in Burmese 11
1.3 Classifiers as derivational markers in Tariana 11
1.4 Inflection and derivation 36
2.1 Spanish 'putting' verbs, differing according to distinctions of Path 91
2.2 The three typological categories for Motion verbs 99
2.3 Acceptable types of causative usage: *die*, *kill*, and *murder* 114
2.4 Lexicalized causation types shifted by grammatical elements 115
2.5 Lexicalization patterns for verbs of posture 120
2.6 Lexicalization patterns for Latin verbs of condition 125
2.7 Derivational patterns for affect verbs focussed on the Stimulus or the Experiencer 135
2.8 Affect verbs in English 136
2.9 Derivation of Experiencer-subject verb roots to Stimulus-subject in Atsugewi 137
2.10 'Cognitive' verbs 138
2.11 Satellites as verb prefixes in German, Latin, and Russian 140
2.12 Typology of Motion verbs and their satellites 154
2.13 Atsugewi aspect satellites' meanings 158
3.1 Dumi nonpast verb inflection (selection) 185
3.2 Typology of positions and formatives 198
3.3 Latin noun paradigms 202
3.4 Chechen noun paradigms 203
3.5 Typology of inflectional classes 204
3.6 Belhare verb paradigm (selection) 204
3.7 Verb paradigms in Latin and Polish 205
3.8 Latin noun paradigm (singular only) 206
3.9 Anêm possessed noun paradigm (selection) (Thurston (1982:37)) 207
3.10 Chechen deictic prefixes 209

3.11 Warrgamay (Pama-Nyungan, Australia; Dixon (1980:287, 329)) 210
3.12 Russian noun paradigm 212
3.13 Abkhaz verb agreement 216
3.14 Belhare intransitive verb agreement of selected tense/aspect/mood forms 217
3.15 So pronouns 221
3.16 Belhare intransitive verb agreement 221
3.17 Hocak (a.k.a Winnebago; Siouan) subject agreement 222
3.18 Rembarrnga pronouns 222
3.19 Old Church Slavic number paradigm 227
3.20 The Chechen verb 'drive' 228
3.21 Behaviour of words and formatives with regard to assignment, spreading, and stacking 238
4.1 Agreement patterns in Russian 250
4.2 Interaction of genders in Paumarí 257
4.3 Gender assignment in Godoberi 259
4.4 Genders III and IV in Archi 260
4.5 Gender assignment in Russian (semantic criteria only) 262
4.6 Examples from the semantic residue in Russian 262
4.7 Noun paradigms in Russian 263
5.1 Use of imperfectives in some languages 302
5.2 Cardinal aspectual operators 304
5.3 Cardinal temporal operators 315
5.4 Cardinal modal operators 329

Contributors

ALEXANDRA Y. AIKHENVALD, La Trobe University

LEONARD TALMY, University at Buffalo

BALTHASAR BICKEL, University of Leipzig

JOHANNA NICHOLS, University of California, Berkeley

GREVILLE G. CORBETT, University of Surrey

ALAN TIMBERLAKE, University of California, Berkeley

BERNARD COMRIE, Max Planck Institute for Evolutionary Anthropology, Leipzig, and University of California, Santa Barbara

SANDRA A. THOMPSON, University of California, Santa Barbara

Acknowledgements

Language typology studies what the languages of the world are like. When people ask 'What is linguistics?', from my point of view one of the best answers is 'the study of what the languages of the world are like'. I an honoured to have been joined by some excellent linguists in the achievement of this second edition of *Language Typology and Syntactic Description* for Cambridge University Press.

I am especially grateful to Matthew Dryer for coming in as co-editor when my health began to fail. Many thanks also to Lea Brown, for the invaluable help she gave Matthew in preparing the manuscript.

The Australian National University has always been generous in its support of my work. Except for the two and a half years I lived in Cairns, 2001 to 2003, it has been my base since I moved to Australia in 1975. I recognize the support I received from James Cook University during my time in Cairns.

I came up with the idea used to organize the first edition at a conference on field work questionaires held at the Center for Applied Linguistics, Washington, DC. I said the best way to prepare for field work is to gain a good idea of what to look for. People thought this was right so I was asked to do the organizing. There have been surveys in the past but I believe none with this scope. The first edition has served as a reference manual and a textbook around the world and I have no doubt the second edition will as well. I have been pleased by the number of good linguists who have told me they have referred to our survey while doing field work valuable to us all.

Interest in the question of what the languages of the world are like is a longstanding one, but in the modern era Joseph Greenberg is an outstanding scholar who did important early work himself and was a model for others to do the same.

In an obituary for Joseph Greenberg by Steve Miller the distinction is made between taxonomists who are lumpers and splitters. Steve Miller says:

> It is fitting that it was Darwin who first thought of the distinction between lumpers and splitters; the OED gives him the first citation of the words as applied to taxonomists. Lumpers gloss over or explain differences in pursuit of hidden unities; splitters do the opposite, stressing diversity.

Joseph Greenberg was a linguistic lumper and his dream of recreating the ur-language of humanity must stand as one of the greatest lumping dreams of all time. He dreamed of deep unity, and he spent an extremely long career pursuing evidence for it. He was still publishing highly technical evidence when he died, at age 85.

It is sad that he never published a manifesto, but he was a scientist and his inductive sensibility was not prone to making sweeping statements unsupported by minute attention to evidence. The nearest he came was in his conclusion to the controversial 1987 Language in the Americas, a book that grouped all languages in the western hemisphere into three families: 'The ultimate goal is a comprehensive classification of what is very likely a single language family. The implications of such a classification for the origin and history of our species would, of course, be very great.' Very great, as in, language was invented once and we might even have some ideas about what that language sounded like.

I was with Joseph Greenberg at Stanford University when he was doing his work, scouring through the part of the library that had grammars, making his counts: if you find construction *x* in a language you will always find, or you will be likely to find, construction *y*. This kind of commonality intrigued him. More from Steve Miller:

The splitters of linguistics have this problem: they're just not as interesting as the lumpers. The splitters' story is that the origins of language are irretrievable, so we should value every language for its expressive ability, but not for its place in the grand drama of linguistic diffusion. Greenberg, and the Nostraticists, and others who have tried to talk about language as a unity, dreamed something that may never be provable, but will continue to inspire us as a story that unites the human race as part of an ongoing story.

We give aid to both the lumpers and the splitters but I believe most of all to the lumpers. Languages differ from each other but only to a certain degree. Humankind is united in its use of language. This is an important message for us all as we go about our pursuits and combine with others to deal with the world.

TIMOTHY SHOPEN

Canberra, Australia
September 2004

Abbreviations and symbols

The following are abbreviations for grammatical terms used frequently in the glosses for examples. Other abbreviations are explained as they are presented.

A	agent (in chapter 2)
A	subject of a transitive verb
A (followed by	absolutive agreement marker (in chapter 3)
ABnumeral, e.g. A3)	abessive
ABL	ablative
ABL (PRIOR)	ablative case in agreement with past tense of verb (in chapter 3)
ABS	absolutive
ACC	accusative
ACT	actual mood
ADJ	adjectivizer
ADL	adlative
AFF	affix
AGT	agentive
ALL	allative
ANT	anterior
AOR	aorist
APPL	applicative
APR	apprehensive
ART	article
ASP	aspect
AUG	augmentative
AUGM	augmented
AUX	auxiliary
CAUS	causative
CISLOC	cislocative
CL	classifier

CL	class (in chapter 7)
CND	conditional
CNTMPL	contemplative
COM	comitative case
COMP	complementizer
COMP	compounding (in chapter 1)
COMPL	completive aspect
COND	conditional
CONJ	conjunct mode (in chapter 3)
CONT	continuous
COP	copula
CSN	comparison
CV	epenthetic syllable
D.O.	direct object
DAT	dative
DECL	declarative
DEF	definite
DEM	demonstrative
DENOM	denominal
DER	derivational
DEST	destinative case
DET	determinator (in Cree verb forms, in chapter 3)
DET	determiner
DIM	diminutive
DIR	direct transitive relation
DO	direct object
DTR	detransitivizer
DU	dual
DUR	durative
E	epenthetic (in chapter 1)
E	experiencer (in chapter 2)
E (followed by	ergative agreement marker (in chapter 3)
ELnumeral, e.g. E3)	elative
EMPH	emphatic
EPEN	epenthetic vowel
ERG	ergative
EXCL	exclusive
EZ	ezafe, izafet
F	feminine
F	figure (in chapter 2)
FAM	familiar

FEM	feminine
FIN	finite form
FUT	future
G	ground
GEN	generic (in chapter 1)
GEN	genitive
GENIT	genitive
GER	gerund
HAB	habitual
HON	honorific
HORT	hortative
HUM	human
IF	imperfect
IFV	imperfective
IMP	impersonal
IMP	imperative (in chapter 7)
IMPF	imperfect / imperfective
IMPV	imperative
INAN	inanimate
INCL	inclusive
IND	indicative
INDEF	indefinite
INDIC	indicative
INESS	inessive
INF	infinitive
INFL	inflection
INFR	inferential
INFV	infinitive
INS	instrumental
INSTR	instructive (in chapter 1)
INSTR	instrumental
INSTRC	instructive (in chapter 7)
INSTRM	instrumental
INTENS	intensifier
INV	inverse transitive relation
IO	indirect object marker
IPFV	imperfective
IPV	imperative
IRR	irrealis
ITER	iterative
ITT	iterative
LAT	lative

LINK	linker
LOC	locative
LOG	logophoric pronoun
M	masculine
M/A	mode–aspect
MASC	masculine
MIN	minimal
MOM	momentaneous aspect
MSC	masculine
NCL	noun class
NEG	negative, negation
NEUT	neuter
NF	non-feminine
NOM	nominative
NOMIN	nominalization
NONHON	nonhonorific
NONOBJ	non-object
NP	noun phrase
NPT	nonpast
NSG	nonsingular (neutralizing a dual vs plural contrast)
NTL	neutral
NTR	neuter
NUM	numeral
NZR	nominalizer
O	direct object
OBJ	object
OBJ	object marker (in chapter 3)
OBJ	objective [argument] (in chapter 5)
OBL	oblique
OPT	optative
P	object of transitive verb
P	patient (in chapter 2)
P	person (in chapter 1)
PART	particle
PASS	passive
PAT	patient
PAUC	paucal
PCL	particle
PCP	participle
PEJ	pejorative
PERF	perfect/perfective
PERF	perfect tense (in chapter 3)

PF	perfect
PFV	perfective
PGR	progressive
PI	past imperfective
PL	plural
PNT	potential
POSS	possessive
POSSPRO	possessive pronoun
POT	potential
PP	past participle
PP	past perfective (in chapter 1)
PRES	present
PROG	progressive
PROGR	progressive
PRP	prepositional case
PRS	present
PST	past, preterite
PT	past
PURP	purposive converb, supine
Q	question marker
R.PAST	remote past
RECIP	reciprocal
REFL	reflexive
REL	relative, relativizer
REM	remote
RESTR	restrictive focus (‘only’; ‘just’)
RLS	realis
S	subject of an intransitive verb
S.SET	specific setting
SBJ	subjunctive
SEQ	sequential
SG	singular
SIM	similarity case (‘like’) (in chapter 3)
SIM	simultaneous (in chapter 5)
SS	same subject
STV	stative
SUB	subjunctive
SUBJ	subject
SUBORD	subordinate
TEL	telic
TNS	tense
TOP	topic

TR	translative (in chapter 1)
TRANSLOC	translocative (locative prefix)
V	verb (root)
VBZR	verbalizer
VCL	verbal classifier
VN	verbal noun
VOL	volitional
WP	witnessed past
1	first person
2	second person
3	third person
4	fourth (obviative) person
1SG	first person singular (etc.)
3PL	third person plural (etc.)
.	separates elements of interlinear that correspond to a single morpheme in the original
ϕ	zero marking
-	affix boundary
=	clitic boundary
(M), (F), ETC.	gender (masculine, feminine, etc.) of noun in chapter 3. (Gender as agreement category is not in parentheses.)
Σ	first element of bipartite verb stem
Σ2	stem alternate
	syllable (annotates left bracket in prosodic transcriptions)
[ ]	glosses in square brackets are zero-marked (in chapter 3)

Roman numerals refer to gender classes.

1 Typological distinctions in word-formation

Alexandra Y. Aikhenvald

0 Introduction

This chapter deals with patterns of word-formation, their classification and parameters of cross-linguistic variation. Grammatical words (section 1) in most languages have an internal structure; the typological parameters which account for their cross-linguistic variation are discussed in section 2. Word-formation processes correlate with syntax in different ways depending on language type. One such word-formation process – known as 'the most nearly syntactic of all' (Mithun (1984)) – is noun incorporation, discussed in section 3.

The structure of words in a language can be more or less iconically motivated (see section 4). Word-formation, traditionally, falls into compounding and derivation. A compound consists of morphemes which could be free (see section 5), while derivation involves the use of different classes of bound morphemes and of morphological processes to form words (see section 6). Word-formation processes vary in terms of their productivity – see section 7. Word-formation processes are prone to distinct patterns of grammaticalization and lexicalization – see section 8. A brief summary is given in section 9, and in section 10 I provide suggestions for field workers describing word-formation in previously undocumented or poorly documented languages.

1 The word

Word-formation accounts for the structured organization of the lexicon. The lexicon is usually conceived of as a list of the form–meaning correspondences conventionalized by speakers, but which are largely arbitrary. However, this list may be structurally organized. The principal function of word-formation is the enrichment of the lexicon by forming new words; for instance, *redden* and *reddish* in English are regular derivations based on *red*.

What is a word? 'Word' has, for a long time, been recognized as a universal unit by scholars of varied persuasions. The concept of the word is, however, at least twofold. Many languages make a distinction between *phonological* and *grammatical* word (though the majority of grammars do not pay enough

attention to this distinction: see Dixon (1977, 1988); Foley (1991); S. R. Anderson (1985a)).

A *phonological* word can be defined as a prosodic unit not smaller than a syllable. Cross-linguistic criteria used to distinguish the phonological word include: (i) stress and other prosodic characteristics; (ii) phonotactics, and phonological rules which apply either word-internally or across word boundaries. See further discussion in Dixon and Aikhenvald (2002).

A *grammatical* word consists of a number of grammatical elements which (i) always occur together, rather than scattered through the clause (the criterion of cohesiveness); (ii) occur in fixed order; and (iii) have a conventionalized coherence and meaning (Dixon and Aikhenvald (2002); see also Dixon (1977:88, 1988:21–31); Matthews (1991)). Criterion (iii) relates to both the number of morphemes per word and the expression of grammatical categories which are obligatory for a grammatical word to be well-formed in a given language. In most non-isolating languages (see section 2), a grammatical word must include at least one inflectional morpheme. For instance, in Yidiny it can have only one (Dixon (1977)). In North Arawak languages of South America a grammatical word must contain at least one root morpheme and not more than one prefix. The presence of inflectional morphemes is not obligatory in grammatical words in Kaingang (Gê), which shows a general tendency toward isolating typology (Wiesemann (1972)).

Grammatical and phonological words often, but not always, coincide (e.g. Lehiste (1964); Dixon and Aikhenvald (2002)). Thus, many languages have clitics which constitute grammatical words on their own but must be attached to another grammatical word within one phonological word and thus cannot form a phonological word on their own, e.g. *-n't* as in English *mustn't*.

Further distinctions within the concept of word include word as an orthographic unit (a useful tool for counting the number of words while composing a telegram; however, it is applicable only to languages with an institutionalized writing system) and word as a lexical unit – that is, a unit which can be treated as one entry in a dictionary (see Mugdan (1994:2551)). Lexical units, whose form–meaning association is hardly predictable on the basis of the meaning of their components, are not limited to a list of words only. Often, a combination of words – a phrase, or even a sentence – can be *idiomatic*, or non-compositional. In English, expressions like *she spilt the beans* or *willy-nilly* ought to be included in lexical listings, based on the arbitrariness of lexical information.

In this chapter, we will limit ourselves only to words as *grammatical* units, concentrating on discovering the principles of the internal structure of words and their cross-linguistic variability, rather than on the arbitrariness of the form–meaning correlations. For this reason idiomatic combinations of words will not be discussed any further. Throughout the chapter, when we say 'word', we are

referring to 'grammatical word' (see Dixon and Aikhenvald (2002), for further discussion).

2 Morphological typology and word-formation

The traditional parameters used for morphological typology of languages starting from the nineteenth century were largely based on the differences in their internal word structure. These parameters are of two kinds. The first one is based on the transparency of morphological boundaries between the morphemes within a grammatical word, and the second one relates to the degree of internal complexity of words (see E. Sapir (1921)).

2.1 Transparency of word-internal boundaries

Based on this parameter, three types of language are recognized: *isolating*, *agglutinating*, and *fusional*.

An *isolating language* typically has a one-to-one correspondence between a morpheme and a word; that is, in such a language every morpheme is an independent word. An example of an almost perfectly isolating language is Vietnamese, as illustrated in (1) (Thompson (1987:207)).

(1) Chị̣ ấy quên
s/he ANAPHORIC forget
'She (or he) forgets', or 'She (or he) has forgotten', or
'She (or he) will forget'

Every word in this sentence is invariable. There is no morphological variation for tense, or for grammatical function. Where English grammar would require a reference to time in the verb in every sentence, in speaking Vietnamese one is not required to have this. The time reference is understood from the context; so (1) could also be translated as 'She (or he) has forgotten' or as 'She (or he) will forget'. If time reference is important, a time word or an aspect marker – also a separate word – can be inserted. In (2), an 'anterior' aspect marker is used in the same sentence as (1) to indicate that the action of 'forgetting' started before the time of the utterance.

(2) Chị̣ ấy đã quên
s/he ANAPHORIC ANTERIOR forget
'She (or he) forgot' or 'She (or he) has forgotten'

It is in general true that every word in Vietnamese consists of just one morpheme; however, the existence of productive compounding and its lexicalization results in the creation of words of more complicated structure, e.g.

hôm nay (day now) 'today', *hôm kia* (day that) 'day before yesterday', *hôm kía* (day that; more remote than *kia*) 'two days before yesterday'.

In an *agglutinating language*, a word may consist of several morphemes but the boundaries between them are clearcut. There is typically a one-to-one correspondence between a morpheme and its meaning, and a morpheme has an invariant shape which makes it easy to identify. Hungarian and Turkish are classic examples. A noun is easily segmentable into a lexical stem, a number affix and a case affix. An extract from the Hungarian noun declension paradigm for *ember* 'man' is illustrated below.

	Singular	Plural
Nominative	*ember*	*ember-ek*
Accusative	*ember-et*	*ember-ek-et*
Dative	*ember-nek*	*ember-ek-nek*
Locative	*ember-ben*	*ember-ek-ben*

In *fusional* – sometimes misleadingly called (in)flectional – *languages* there is no clear boundary between morphemes, and thus semantically distinct features are usually merged in a single bound form or in closely united bound forms. Extracts from Russian nominal paradigms for *dom* 'house' and *koška* 'cat' below illustrate this point.

	Declension 1		Declension 2	
	Singular	Plural	Singular	Plural
Nominative	*dom*	*dom-a*	*košk-a*	*košk-i*
Accusative	*dom*	*dom-a*	*košk-u*	*košek*
Dative	*dom-u*	*dom-am*	*košk-e*	*košk-am*
Instrumental	*dom-om*	*dom-ami*	*košk-oj*	*košk-ami*

An affix like *-ami* cannot be segmented into a marker for number and a marker for case; and in a word like *košek* ('cats' accusative plural) the stem itself is fused with case and number. Along similar lines, in Latin the final *-a* of *femina* 'woman' expresses the meanings: nominative case, singular number and feminine gender (as well as first declension).

The term *(in)flectional*, sometimes used in place of fusional, is misleading: we will see in section 11 that both fusional and agglutinating languages, as opposed to isolating languages, can have inflectional morphology.

Fusion and agglutination are best treated as quantitative notions. Even the 'classic' agglutinating languages such as Turkish or Hungarian may be problematic with respect to the treatment of boundaries and the existence of variants of morphemes (allomorphs). These languages are known for vowel harmony across morphemic boundaries, e.g. Hungarian *ember-ek-ben* (man-PL-LOC) 'in men', but *ásztal-ok-ban* (table-PL-LOC) 'in tables'. In addition, Hungarian has a

certain amount of stem alternation in the formation of plurals (e.g. *szó* 'word', pl. *szá*v-*a*-*k)* (see Hagège (1990) on the tendency of an agglutinating morphology to develop into a fusional, or partly fusional, type). Various phonological processes apply across morpheme boundaries, and, as a consequence, the morpheme boundaries may become blurred, which yields the creation of fusional morphology (see section 6).[1]

2.2 *Internal complexity of grammatical words*

The second typological parameter has to do with the number of morphemes per word. This typological dimension is largely complementary to that described in section 2.1.

Analytic languages tend to have a one-to-one correspondence between a word and a morpheme; they have few if any bound morphemes. Vietnamese (1–2 above) or Mandarin Chinese are good examples of analytic languages.

In contrast, in *synthetic* languages a word consists of several morphemes, and there are numerous bound morphemes. Hungarian or Russian are representative of synthetic languages.

Polysynthetic languages (also sometimes called 'incorporating': see section 3, on the reasons for distinguishing these terms) are characterized by extreme internal complexity of grammatical words. Here, the bound morphemes often express semantic content reserved for lexemes in languages of other types. Polysynthesis basically refers to the possibility of combining large numbers of morphemes (lexical and grammatical) within one word, as in the following example from West Greenlandic (Fortescue (1994:2602)):

(3) anigu-ga-ssa-a-junna-a-ngajal-luinnar-simassa-galuar-put
avoid-PASS-PART-FUT-be-no.longer-almost-really-must-however-3PL.INDIC
'They must really almost have become unavoidable but . . .'

Interest in polysynthesis has grown considerably since the 1990s, due to an increasing amount of new data from different parts of the world (Foley (1986, 1991); De Reuse (1994); Fortescue (1994); among others). The following traits tend to cluster in polysynthetic languages, although none of them is defining by itself (Fortescue (1994:2601)):

(i) noun stem incorporation within the verbal complex, and incorporation of adjectival stems within nouns (see section 3);

[1] E. Sapir (1921) suggested a fourth type: *symbolic* languages. These languages utilize internal changes, such as ablaut, vowel and consonant changes, and changes in stress and tone, as a means of marking grammatical contrasts. This type has never been as widely used in typological classification of languages as the others, mainly because these internal changes are also widely used in fusional languages, and it is hard to draw a boundary.

(II) a large inventory of bound morphemes, together with a limited set of independent stems;
(III) derivational processes productive in the formation of individual sentences, the verbal word being a minimal sentence;
(IV) pronominal cross-referencing of subjects, objects, and sometimes also of other arguments (obliques, or datives) on the verb, and of possessors on nominal forms;
(V) integration of locational, instrumental and other adverbial elements (manner, etc.) into the verb complex as bound morphemes;
(VI) many possible affixal 'slots', just a few of them obligatory, within a verbal word.

Concomitant properties of polysynthetic languages include relatively free pragmatic constituent order, possibilities of variable morpheme ordering and head-marking.

Many, but not all, polysynthetic languages have noun incorporation (section 3). Most can have a wide range of recursively occurring affix types (verbalizers, nominalizers, adverbial type 'postverbs') with an extremely large overall stock of affixes (e.g. 400–500 in West Greenlandic, and 200 in Kwakwala). Yet other languages are typified by a large number of affixes attached to different slots only within a verbal complex ('field-affixing': Fortescue (1994:2602)). They can be suffixing (Yupik, or West Greenlandic), or suffixing and prefixing (e.g. Nadëb, from the Makú family; Guahibo languages from Colombia; or North Australian languages).

The combination of these properties is also attested. A combination of incorporation and 'field'-affixing can be illustrated with the structure of the verb complex in Traditional and Modern Tiwi (Osborne (1974); Lee (1987: 152–3)) – see table 1.1. (Modern Tiwi, spoken by the younger generation, has been simplified within a contact situation: Lee (1987:155–6).)

Example (4) shows a chain of prefixes in Traditional Tiwi. All these prefixes are said to be obligatorily used.

(4) warta a-watu-wuji-ngi-mangi-rr-akupuraji yiripuwarta
bush 3SG.MASC-morning-CONT-CV-water-CV-fall high.tide
'The high tide is falling [literally 'water-falling'] [exposing the] land (bush)' (Jennifer Lee, p.c.)

Historically, polysynthetic morphology often arises from the combination and subsequent grammaticalization of independent roots. Thus, Fortescue (1992) suggests 'that contemporary Eskimo languages may have developed their complex morphophonemic patterns from a more agglutinative pre-Proto-Eskimo stage' (cf. also Foley (1997) for Yimas; Aikhenvald (2003) for Tariana).

Since polysynthetic structures are most often found in head-marking languages, Nichols (1986) suggested that there are no polysynthetic nouns.

Table 1.1 *Morpheme slots in Tiwi verb (Lee (1987:152–5))*

Traditional Tiwi	Modern Tiwi
1. Subject	yes
2. Tense: past, non-past	yes
3. Locative: distant, directional, distant in time	yes
4. Mood 1: subjunctive, frustrative	yes
5. Mood 2: irrealis	yes
6. Temporal 1: 'in the morning'	no
7. Direct object or indirect object	no
8. Aspect 1: durative or non-past habitual, inceptive, common activity	yes
9. Stance: away from camp, or distant in time; walking along	no
10. Emphatic	yes
11. Connective	yes
12. Temporal 2: 'in the evening'	no
13. Concomitative	no
14. ± 1 or 2 incorporated forms	no
15. Verbal root	yes
16. Voice: causative, completive, reflexive, reciprocal	yes
17. Aspect 2: movement; 'on the way'	yes
18. Aspect 3: repetitive, past habitual	yes
19. Locative	no

However, nouns in some Australian languages (Dench and Evans (1988)) and in some languages from South America (Aikhenvald (1999c)) have been shown to be inflectionally polysynthetic, since they have multiple marking of grammatical function known as 'double case' (see also Plank (1995)).

The distinction between analytic and synthetic languages is a continuum rather than a dichotomy, since languages display different degrees of synthesis. The degree of synthesis or analysis in a given language can be calculated, for instance, by dividing the number of morphemes in a sentence by the number of words. Some languages are considered more synthetic than others. Linguists often talk about 'mildly' polysynthetic languages. This is reflected in the approach of Greenberg (1954) who suggested the use of a quantitative index, M(orpheme) per W(ord) to calculate the degree of synthesis in a language. See Comrie (1981a:44–5) for further discussion of problems which arise there.

Languages which can be considered almost entirely analytic are the isolating languages of Southeast Asia – e.g. Mandarin Chinese, Classical Chinese and Vietnamese – and of West Africa – e.g. Igbo. The languages of Europe, Asia and North Africa are predominantly synthetic, while polysynthetic languages are concentrated in North and South America, in Siberia, in the north of Australia and in some parts of Papua New Guinea (Foley (1986)).

2.3 Integrating the two parameters

The degree of synthesis and the treatment of morphological boundaries are relatively independent typological parameters. For a description of a previously undocumented language, it is not enough to say that it is 'analytic', or that it is 'isolating'. It is true that isolating languages tend to be analytic, but the reverse would be wrong: English, which has some fusional morphology, makes extensive use of analytic constructions.

Polysynthetic languages are often agglutinative in that the morpheme boundaries are clearcut, and there is little allomorphic variation. However, some polysynthetic languages do have elements of fusion. For instance, Greenlandic has a well-developed array of fused portmanteau inflections with a great morphophonemic complexity – see Fortescue (1992). The fusion of morphemes in a polysynthetic language is illustrated by (5), from Chiricahua Apache, an Athabascan language (Hoijer (1945:15)). Fused morphemes are underlined.

(5) hà-ń-ʔàh
out.of-2SUBJ+IMPF-handle.a.round.object+IMPF
'you take a round object (out of enclosed space)'

The degree of morpheme fusion and of synthesis have to be defined independently of one another. Figure 1.1 illustrates how the two can be plotted together. Examples of languages are given underneath the diagram.

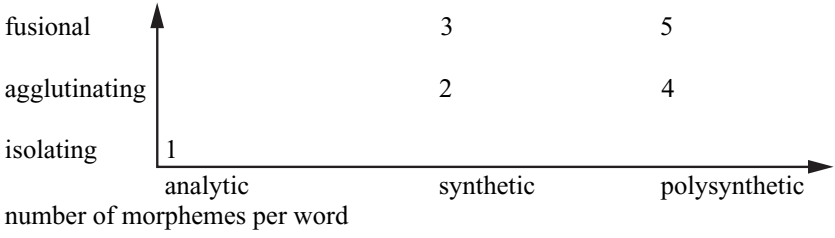

Figure 1.1 Interaction of two types of parameters in word-formation.

(1) Vietnamese and Classical Chinese are typical examples of isolating analytic languages.
(2) Hungarian is a typical agglutinating synthetic language.
(3) Russian is a fusional synthetic language.
(4) Yupik Eskimo is a polysynthetic agglutinating language.
(5) Chiricahua Apache is a polysynthetic fusional language.

2.4 Word-formation and syntax in languages of different types

The two sets of parameters illustrated in Figure 1.1 correlate with other properties. Isolating analytic languages tend not to have obligatory grammatical categories ordinarily shown in fusional or agglutinating languages, such as tense and case or agreement in gender or number (see examples (1–2) from Vietnamese).

As we will see in the following sections, compounding is widespread in isolating languages, while derivation is a property of languages of other types; this follows from the tendency to have a one-to-one correspondence between a morpheme and a word in isolating languages.

Analytic languages employ periphrastic constructions in syntax whereas synthetic languages tend to express similar meanings within an individual word by means of its affixes.

In Japanese, a synthetic language, passive – whereby the object of a transitive verb becomes the subject of an intransitivized verb and the original subject of the erstwhile transitive verb gets demoted – is expressed with an affix, as in (7). Example (6) is the underlying transitive clause.

(6) Naomi-ga Seiji-o ut-ta
Naomi-SUBJ Seiji-O hit-PAST
'Naomi hit Seiji'

(7) Seiji-ga Naomi-ni ut-are-ta
Seiji-SUBJ Naomi-by hit-PASS-PAST
'Seiji was hit by Naomi'

In contrast, an analytic language, such as Vietnamese, typically employs a periphrastic passive construction, as illustrated in (9), the passive of (8).

(8) thây pha tôi
teacher punish I
'The teacher punishes me'

(9) tôi bi thây pha
I suffer teacher punish
'I am punished by the teacher'

English, also a fairly analytic language, tends to employ periphrastic constructions which correspond to affixal constructions in more synthetic languages. Examples (10) and (11) illustrate an active and a passive sentence, respectively, in Latin; translations show their English counterparts.

(10) Mulier hominem videt
woman man+ACC.SG see+PRES+3SG
'The woman sees the man'

(11) Homo ā muliere vidētur
man by woman+ABL.SG see+PASS+PRES+3SG
'The man is seen by the woman'

Analytic isolating languages, such as Mandarin Chinese, tend to have no marking of grammatical relations other than constituent order (whereby 'the actor of a verb, if expressed, must precede the verb': LaPolla (1995:297)). Compare (12) and (13).

(12) wǒ men tjɛɛn tsin
I PL play piano
'We are playing the piano' (or 'we are playing the pianos', 'we are going to play the piano', etc.)

(13) ta da wǒ men
s/he hit I PL
'She or he is hitting us', 'she or he will hit us', etc.

Since the overt noun phrases are often omitted, the participants have to be inferred from the context. Thus, isolating languages are heavily context-dependent; it has been argued that in Chinese there has been no grammaticalization of the syntactic relations 'subject' and 'object' (see LaPolla (1995), for further discussion).

Numeral classifiers as independent words tend to occur in analytic isolating languages (Aikhenvald (2000)). A numeral classifier is illustrated in (14), from Hmong, a Hmong-Mien language from China (see Bisang (1993); Jaisser (1987:172)):

(14) Lawv muaj rau tus me nyuam
they have six NUM.CL:LIVING.BEING child
'They have six children'

When inanimate nouns appear with different classifiers, these highlight different aspects of their meaning. A well-known example from Burmese (Becker (1975:113)) illustrates this point. 'River' can be spoken of in at least eight contexts, shown in table 1.2. Numeral classifiers here are comparable to derivational affixes in more synthetic languages. The specific classifiers can thus add information about the referent, since they allow speakers to distinguish one sense of the referent from all the others. The 'repeater' classifier *myi?* – identical to the noun itself – in table 1.2 indicates that a river is looked upon just as a river, and

Table 1.2 *Reclassification of an inanimate noun in Burmese*

noun	numeral	classifier	translation
myiʔ	*tə*	*yaʔ*	'river one place' (e.g. destination for a picnic)
myiʔ	*tə*	*tan*	'river one line' (e.g. on a map)
myiʔ	*tə*	*hmwa*	'river one section' (e.g. a fishing area)
myiʔ	*tə*	*ʻsin*	'river one distant arc' (e.g. a path to the sea)
myiʔ	*tə*	*thwɛ*	'river one connection' (e.g. linking two villages)
myiʔ	*tə*	*ƥa*	'river one sacred object' (e.g. in mythology)
myiʔ	*tə*	*khuˈ*	'river one conceptual unit' (e.g. in a discussion of rivers in general)
myiʔ	*tə*	*myiʔ*	'river one river' (the unmarked case)

Table 1.3 *Classifiers as derivational markers in Tariana*

pa-da	*episi-da*	'one motor' (one round metal thing)
one-CL:ROUND	metal-CL:ROUND	
pa-kha	*episi-kha*	'one metal wire'
one-CL:ROPE.LIKE	metal-CL:ROPE.LIKE	
pa-pukwi	*episi-pukwi*	'one metal ring'
one-CL:ROUND.HOLLOW	metal-CL:ROUND.HOLLOW	

helps discard other senses (see further examples and discussion in Aikhenvald (2000:ch. 12)).

In synthetic languages numeral classifiers tend to be affixes. In some, such as Tariana, a North Arawak language from northern Brazil, affixed numeral classifiers can be attached to nouns themselves to form new words, as shown in the examples in table 1.3.

That is, analytic and synthetic languages employ different techniques to achieve the same end – enriching their lexicon. While synthetic languages rely on the internal structure of their grammatical words, analytic languages employ syntactic devices.

3 Noun incorporation

The term *noun incorporation* refers to morphological structures in which a nominal constituent is added to a verbal root, and the resulting construction is both a verb and a single word. Incorporation serves to derive lexical items. This process also has morphological, syntactic and discourse consequences, since it creates structures that often affect syntactic relations within a clause and have pragmatic functions in discourse. Incorporation is a morphological

process which brings word formation and syntax close together (see Mithun (1984)).

Incorporating languages are erroneously equated with polysynthetic languages. As was shown in section 2.2, polysynthetic languages do not always have incorporation. And languages with incorporation need not be polysynthetic – this is the case with numerous Austronesian languages such as Fijian or Mokilese. See Kroeber (1911), E. Sapir (1911), Sadock (1980), De Reuse (1994) and especially Mithun (1984, 1986, 1994) for detailed and illuminating accounts of incorporation.[2]

3.1 Formal properties of incorporation

Incorporating structures can be classified according to what type of material gets incorporated (section 3.1.1), and the degree of formal cohesion between the components (section 3.1.2).

3.1.1 What material gets incorporated

The incorporated nominal constituent can consist of (I) a free form of a noun, (II) a bare noun root, (III) a special suppletive or semisuppletive form, or (IV) a whole noun phrase.

3.1.1.1 (I)Incorporation of a free form of a noun. In many languages the incorporated noun does not undergo any changes, as in (16), from Nadëb, a South American language from the Makú family (Weir (1990:323ff.)) where the noun 'house' gets incorporated, as compared to (15), where the same noun occurs on its own.

(15) Subih tɔb tɨ̃ɨh ta-ma
Subih house 1SG THEME-make
'I am making Subih's house'

(16) Subih tɨ̃ɨh tɔb-ta-ma
Subih I house-THEME-make
'I am making a house for Subih'
(literally 'I am house-making Subih')

3.1.1.2 (II)Incorporation of a bare noun root. This is also a frequent type. Example (18), from Ngan.gityemerri, an Australian language (Reid (1990:190)), is an incorporated version of (17). The incorporated noun, 'leg', has been stripped of its noun class prefix *da-* which can be seen in (17).

[2] Unmotivated extensions of this term to various kinds of derivations abound in Baker (1988, 1995).

(17) ngudeny-fityi da-garri
1SG.SUBJ.PERF.shove.DTR-roll NCL-leg
'I crossed my legs'

(18) ngudeny-garri-fityi
1SG.SUBJ.PERF.shove.DTR-leg-roll
'I crossed my legs'
(literally 'I am cross-legged')

3.1.1.3 (III)Incorporation of a suppletive or reduced stem. Special suppletive or semisuppletive stems, distinct from the noun root, are found in a few Northern Australian languages. Example (19), from Traditional Tiwi, illustrates the incorporation of two constituents (Lee (1987:164)). The free form for incorporated *kiji* 'stick' is *taŋini* (Osborne (1974:49)), and that for *maripi* 'chest' is *pipwa* (1974:50).

(19) nga-mpi-ri-kiji-maripi-rrituwa
we(INCL)-NP:her(DO)-CV-STICK-CHEST-slit.open
'We slit the chest [of a goose: FEM] with a stick'
(literally 'we chest-stick-slit-her')

An incorporated form can be a truncated version of the free-form noun. In Murrinh-Patha (Australian; Walsh (1996); Knight (1993:43)), the free form *lamala* 'shoulder' is incorporated as *mala*, and *nginipunh* 'internal body' as *ngini*. Palikur, a polysynthetic Arawak language from Brazil and French Guiana, has a closed set of incorporated body parts. Some of them coincide with the full noun, e.g. *duk* 'chest', and some get shortened and undergo idiosyncratic changes, e.g. free *kugku*, incorporated *kug(a)* 'foot'; free *utyak*, incorporated *-(h)ot(a)* 'eye' (Aikhenvald and Green (1998)).

3.1.1.4 (IV)Incorporation of the whole NP. This type of incorporation is not at all frequent. Incorporation usually takes place when the head noun of a noun phrase is neither specific nor referential. Example (20), from Boumaa Fijian (Dixon (1988:227)), illustrates how a whole possessive noun phrase – '*e-dra-i'a* meaning 'their fish' – can get incorporated (cf. (33), from Nahuatl).

(20) saqa.-['e-dra-i'a]
cook-[their-edible.thing-fish]
'[they will return home and] cook their fish'

Example (21), also from Boumaa Fijian, shows incorporation of an attributive noun phrase consisting of a noun and an adjective.

(21) ‘ana-[waci-po’i]
eat-[cooked.taro.leaves-rolled]
‘eat rolled taro leaves’

In Fijian it is even possible to incorporate a noun phrase with an ‘or’ disjunction.

(22) e la’i taa-[niu-se-bu’a]
ASP go chop-[copra-or-firewood]
‘He’s gone to chop copra or firewood’

A comitative noun phrase can be incorporated in Rembarrnga, an Australian language (McKay (1975:171)), as in (23).

(23) ŋa-[parta-winta]-rtuŋʔ-miñ
1SG.S-[spear-COMITATIVE]-fall-PUNCTUAL
‘I fell with a spear [sticking out of me]’

3.1.2 The degree of formal cohesion between components

There are two possibilities. A verb and a nominal constituent can be juxtaposed, but remain separate phonological words, as in numerous Austronesian languages (see Mithun (1984:849–50)). Example (24), from Boumaa Fijian (Dixon (1988:227)), is a transitive sentence where the object noun phrase refers to some specific breadfruit. Example (25) contains an incorporated noun phrase which is an independent phonological word. Unlike (24), (25) is an intransitive sentence referring to a generalized activity of ‘breadfruit-eating’ rather than eating any particular breadfruit. The incorporated noun has lost its syntactic status as an argument of the verb (direct object) and it cannot be modified with an article, or have specific reference.

(24) e ˈani-a a uto
3SG.A eat-3SG.O ART breadfruit
‘He is eating the/some breadfruit’

(25) e ˈana-uto
3SG.A eat-breadfruit
‘He is eating breadfruit’
(literally ‘is engaged in breadfruit-eating’)

Alternatively, the formal cohesion between the incorporated noun and the verb can be tighter: they constitute one phonological word and take a single stress, as in Ngan.gityemerri (18), Tiwi (19) and Nadëb (16). In (26), from Cayuga (Iroquoian; Mithun (1994)), the incorporated noun ‘berry’ enters into the syllable count for the purpose of stress assignment (the fourth syllable is stressed), and participates in laryngeal spreading, a phonological process

whereby vowels within one phonological word become laryngealized (shown by the dot underneath every vowel).

(26) kạhyakwạhskḵhḵ:ʔ – surface realization
k-ahy-kw-ahs-kḵhḵ:ʔ – underlying form
1.AGENT-berry-get-HABITUAL-FORMER.PAST
'I used to berry-pick'

3.2 Functional types of incorporation

We can distinguish five functional types of incorporation (roughly following Mithun (1984, 1994)).[3]

3.2.1 Type 1. Lexical compounding

If a language has any noun incorporation at all, it has lexical compounds. Lexical compounding is derivation of a complex lexical item from a combination of two or more stems to refer to a 'name-worthy' unitary activity, such as 'berry-picking' in (26). Lexical compounding often derives intransitive verbs. Consider (27) and (28), from Mokilese, an Austronesian language (Harrison (1976:162)) (cf. also (24) and (25) above). Example (27) contains a specific noun phrase with a determiner: 'these coconuts'.

(27) Ngoah kohkoa oaring-kai
I grind coconut-these
'I am grinding these coconuts'

The noun 'coconut' is incorporated in (28). This sentence refers to a habitual activity of grinding coconuts, and cannot refer to any particular individualized coconuts.

(28) Ngoah ko-oaring
I grind-coconut
'I am coconut-grinding'

Verb-object compounds are extremely productive in Mandarin Chinese, e.g. *jié-hūn* (tie-marriage) 'marry' and *kāi-dāo* (open-knife) 'operate on' (Li and Thomson (1981:75–7)). Verb-subject compounds in Mandarin Chinese involve intransitive adjectival verbs, e.g. *xīn-ruǎn* (heart-be.soft) 'be softhearted', *mìng-kǔ* (life-be.bitter) 'be unfortunate'; only some are action verbs, e.g. *tóu-téng* (head-ache) 'have a headache', *bīng-biàn* (soldier-rebel) 'mutiny', *dì-zhèn* (earth-quake) 'have an earthquake' (1981:71–2). There are hardly any examples of compounding of transitive subjects.

[3] Type 1 here corresponds to Mithun's type I; type 2 and type 3 to her types II and III; type 5 to her type IV (classificatory noun incorporation); type 4 has not been considered in her paper.

The semantics of compounds is often non-compositional, e.g. Korean *kil-tulta* (road-enter) 'get used to'. One of the criteria for verb–object compounds in Mandarin Chinese involves the non-compositionality of their meaning (Li and Thomson (1981:71–2)).

Compounding is typically used for 'naming' some important activity, e.g. *baby-sit* (but not *pig-sit*, unless someone employs someone else to take care of an unusual pet), *fund-raise, home-deliver, problem-solve*, and Boumaa Fijian *unu-wai* 'drink water' (water-drink) (Dixon (1988)). Lexical compounds of this sort may have to be entered into a dictionary as separate lexical items, since their meaning is often non-compositional, e.g. Hungarian *világ-latszani* (world-see) 'travel'; Paumari (Arawá, Brazil) *-va'i-hoki* (liver-be.alive) 'remember'; Mayali (Australian; N. Evans 1991, 1996) *ngei+bu* 'flower+hit' 'to flower' and *danj+bu* (spear+hit) 'to spear'.

To understand the meaning of non-compositional compounds one has to be familiar with the culture. For instance, Boumaa Fijian *unu-tii* (lit.: 'drink tea') is a lexical compound which in fact refers to something more than just tea-drinking: it means a small meal 'in which tea drinking is accompanied by eating bread, scones or pancakes'. A compound *unu-sede* literally means 'drink money (cents)', and 'describes a kava-drinking party where each participant contributes a small sum, perhaps twenty cents, in order to raise money for a specific purpose' (Dixon (1988:227)).

3.2.2 Type 2. The manipulation of case

The incorporation of an argument can have a syntactic effect; then it results in the change of syntactic relations within a clause. Consider (15) and (16) from Nadëb (Makú). As a syntactic position was vacated by the incorporated nominal, 'house', the erstwhile possessor, Subih, gets 'advanced' into the position of the direct object.

The manipulation of case often also has concomitant semantic and pragmatic effects. That is, incorporation permits speakers to cast important participants into core roles – S or O.[4] The semantic difference between (15) and (16) in Nadëb is that in (16) the benefit for Subih is considered more important than building a house. In (29), from Cayuga, the victim of 'scalping', which is 'presumably, of greater overall interest than the scalp' – since the story centres around the human protagonists and not their body parts – can occupy a core case role due to incorporation (Mithun (1994:5025)).

(29) a-t-he-nõh-hk
FACTUAL-DUAL-1SG.AGENT/M.SG.PATIENT-scalp-pick
Literally 'I scalp-picked him', that is, 'I scalped him'

[4] In this chapter I employ the standard abbreviations S (intransitive subject), A (transitive subject) and O (transitive object). Some other writers use P or U instead of O.

3.2.3 Type 3. The regulation of information flow

Incorporation is often used to background known or unimportant information in discourse (see also the Cayuga examples in Mithun (1994:5025)). In Nahuatl (Uto-Aztecan; Merlan (1976)) a new entity is introduced by an independent ('external') noun phrase, as in (30).

(30) askeman ti-'-kwa nakatl
never you-it-eat meat
'You never eat meat'

Once the noun is old information, it is incorporated, as in (31) from the same conversation as (30).

(31) na' ipanima ni-naka-kwa
I always I-meat-eat
'I always eat meat'

The incorporated noun in Nahuatl (Merlan (1976:188)) is unmarked for features such as definiteness or specificity. Only non-incorporated nouns can be contrastive, as in (32); compare the incorporated counterpart in (33).

(32) ni-ki-išmati itos
1SG-it-know 3SG+voice
'I know him by his voice' (and not in some other way)

(33) nitos-išmati
1SG+3SG+voice-know
'I know his voice'

3.2.4 Type 4. Incorporation of modifiers

Incorporation of a modifier is found in some Australian languages. Adjectival modifiers can be incorporated only if the head noun is the subject of an intransitive verb, as in (34), from Rembarrnga (Australian; McKay (1975:290)), or a direct object, as in (35), from Mayali, also Australian (N. Evans (1996:102)).

(34) Ø-kartpurr-mañ
3.MIN.SUBJ-wounded-went
'He [buffalo] went away wounded'

(35) barri-darrgid-ma-ngi
3.AUGM/3-alive-pick.up-PI
'They pick [it, i.e. a crocodile] up alive'

3.2.5 Type 5. Classificatory incorporation

A generic noun can be incorporated to narrow the scope of the verb characterizing its direct object or the intransitive subject. Semantically this is similar to

generic noun classifiers (see Dixon (1982); Aikhenvald (2000:ch. 3)); cf. (36), from Mayali:

(36) ga-rrulk-di an-dubang
3NP-GEN.CL:TREE-stand CLIII-ironwood.tree
'An ironwood tree is there'
(literally 'An ironwood tree tree-stands')
(N. Evans (1996:77))

An incorporated noun can get grammaticalized as a verbal classifier, categorizing the argument (O or S) in terms of its shape (cf. Mithun (1984); Aikhenvald (2000)). Mundurukú, a Tupí language from Brazil (Gonçalves (1987: 42)), has over 100 classifiers which refer to shape and form; most of them come from body-part nouns. In (37), classifier *-ba*^{4} 'long and rigid' refers to O ('banana'); it comes from a body-part term meaning 'arm'.[5]

(37) Be^{3}kit^{2}kit^{2} a^{2}ko^{3}-ba^{4} o'^{3}-su^{2}-ba^{2}-do^{3}bu^{2}xik^{3}
child banana-CL:LONG.RIGID 3SG-POSS-CL:LONG.RIGID-find
'A child found a banana'

One language may have more than one type of noun incorporation. This is an important argument in favour of the proposed typology. Different types of noun incorporation can differ just in their semantics. Retuarã, a West Tucano language from Colombia, has type 1 and type 2 incorporation. If an incorporating structure describes a customary activity, lexical compounding (type 1) is employed, yielding combinations like firewood-feed = make fire; medicine-put = treat; or (38):

(38) kopakaha dã-tã?ãpi-hãã-ti-ko?o
now 3PL-coca-put.it-PERF-PAST
'Now they have chewed-coca'

If the activity is not customary, type 2 incorporation (manipulation of case) occurs as in (39) (Strom (1992:100)). In this example the noun 'seat' is incorporated into the verb 'put', and 'canoe' becomes a direct object: it is cross-referenced on the verb with the prefix *sa-* 'third person inanimate singular object':

(39) bikitoho sa-ki-terĩ-hãã-rãyũ)
morning 3INAN.SG.O-3MASC.SG.A-seat-put-future
'In the morning he will put seats in it (canoe)'
(literally 'he will seat-put it')

According to Mithun (1984), there are hierarchical relations between the types of incorporation. If a language has classificatory noun incorporation

[5] Note that numbers indicate tones, with 1 being high tone and 4 low tone.

(type 5), it will also have incorporation as regulation of information flow (type 3), as well as case manipulation incorporation (type 2) and lexical compounding (type 1). This implicational hierarchy suggests a path for the evolution of noun incorporation. Noun incorporation starts from lexical compounding, and then goes through other types, with classificatory noun incorporation as its latest stage.

3.3 Syntactic functions of incorporated nouns, and their incorporability

Incorporated nouns typically are in S (intransitive subject) or O (direct object) (see Keenan (1984), and examples (16–39)). According to Mithun (1984:875), if a language incorporates nouns in just one function, they will be direct objects; if a language incorporates only two types of arguments, they will be direct objects and subjects of intransitive verbs; many languages also incorporate instruments and also locations. Example (40), from Chipewyan, an Athabascan language, illustrates the incorporation of an instrument, 'hook' (Cook and Wilhelm (1998:59)). Example (41) contains the same noun as a free form. Incorporated forms are used if the action is more habitual, with little specification of the incorporated participant.

(40) na-jə́th-the-Ø-Ø-da
ITER-hook-M/A-3SG-VCL-sit
'S/he is fishing again'
(literally 'sitting with a hook')

(41) jə́th ghą the-Ø-Ø-da
hook with M/A-3SG-VCL-sit
'S/he is fishing'
(literally 'sitting with a hook')

The subject of transitive verbs can hardly ever be incorporated.[6] Alamblak (Sepik Hill, Papuan; Bruce (1984:170)) is unusual in that it permits the incorporation of a body part whose possessor is in A function. Example (42) is a transitive sentence with two unincorporated arguments, 'child' and 'foot'. In (43), the A ('foot') is incorporated into the verb. This is incorporation of type 2, since it includes manipulation of case with semantic and pragmatic consequences: (42) is about the child's foot, and (43) (an intransitive clause) is about the child.

(42) yën-r wura-t yëhne-më-t-r moh-ohat-n
child-3SG.M foot-3SG.F descend-R.PST-3SG.F-3SG.M hole-path-S.SET
'A child('s) foot went down the hole on him'

[6] Verbal classifiers operate similarly; only in a few exceptional cases do they characterize A; see Aikhenvald (2000) on Motuna and Nasioi, Papuan languages from Bougainville; see also Onishi (1994).

(43) yën-r wura-yëhne-më-r moh-ohat-n
child-3SG.M foot-descend-R.PST-3SG.M hole-path-S.SET
'A child went down into the hole (up to his) foot'
(literally 'child foot-descended into the hole')

Different constituents may be incorporated under different conditions. In Alamblak (Bruce (1984)) any noun in S, O or locative function can be incorporated in a dependent clause; while in a main clause only inalienably possessed nouns can be incorporated.

Body parts and relational nouns (e.g. terms like *child-of*) are more likely to be incorporated than nouns of other semantic groups (see Zhivov (1978); see Merlan (1976:188) for a functional explanation). In many languages only body-part nouns can be incorporated (e.g. Australian languages – N. Evans (1996), Walsh (1996) – or Palikur, an Arawak language from Brazil – Aikhenvald and Green (1998)). In most Amazonian languages (Guahibo, Nadëb) only obligatorily possessed nouns can be incorporated.

In most cases, members of closed classes cannot be incorporated. Boumaa Fijian is unusual in allowing lexical incorporation of the interrogative *cava* 'what', as in *unu-cava* 'drink what?' (Dixon (1988:227)). Further restrictions on incorporability of nouns follow from their referential properties. Definite or referential nouns cannot be incorporated. This is the reason why personal names are rarely (if ever) incorporated.

Some languages allow more than one constituent to be incorporated simultaneously; see Walsh (1996:358) on incorporating two body-part terms in Murrinh-Patha, and example (19), from Tiwi. Nadëb allows the incorporation of various constituents with the pragmatic result that the 'new' direct object comes into focus (see (15) and (16)). It is also possible in Nadëb to incorporate two or even three nouns, but this is not common (Weir (1990: 332)). Example (44) illustrates two incorporated nouns:

(44) a hoonh kad tɔb-nooh-ga-jʉtt
2SG+POSS grandmother uncle house-mouth-THEME-close
dʉk
be.suspended
'Uncle closed the door of your grandmother's house'
(literally 'Uncle house-mouth-closed your grandmother';
the effect on the grandmother is emphasized)

Adverbs and adpositions (prepositions or postpositions, depending on the language) can form part of lexical compounds (type 1 incorporation), e.g. English *overdo*, *outdo*, *underrate*. Incorporation of adverbs and adpositions is often used as a valency-changing device (similarly to type 2, manipulation of case). Incorporation of an adposition in Nadëb is functionally

similar to applicative. If the verb is intransitive, the argument of the postposition becomes O, and the original S becomes A. Example (45) is intransitive, and (46) is transitive.

(45) ɛɛ$_S$ a-hɨng [hxɔɔh go]
father FORMATIVE-go.downriver canoe in
'Father goes downriver in a canoe'

(46) hxɔɔh$_O$ ɛɛ$_A$ ga-hɨng
canoe father in-go.downriver
'Father goes downriver in a canoe'
(literally 'Father goes-downriver-in a canoe')

This incorporation has a syntactic effect: an argument of a postposition cannot be relativized, but a direct object can be (see above). It also has a discourse effect: a direct object is more topical than an argument of an adposition.

We have seen that incorporation is a means of enriching the lexicon: lexical compounding serves to create new lexemes. It may also have a syntactic effect, altering grammatical relations within a clause. Its pragmatic effect has to do with highlighting a new participant or backgrounding an old one. Finally, incorporation can also have a stylistic effect: for instance, constructions with incorporation have been described as 'more idiomatic, more elegant' for Carrier, an Athabascan language (Cook and Wilhelm (1998:61)) (see section 7.5).

4 Structure and iconicity in word-formation

The notion of structure in word-formation implies that some items in the lexicon can be considered partially motivated in terms of an association between their form and their meaning. Some words in a language are 'unanalysable'; the association between form and meaning is conventionalized by speakers' usage. Other words consist of isolable parts with form and meaning of their own combined in a principled way.

Languages differ in how much derivational motivation (and hence derivational complexity) they allow for individual words. For instance, the body-part terms *eye*, *beard* or *moustache* in English are not decomposable; the association between their phonological form and their meanings can be considered arbitrary. In contrast, the word *eye-lash* consists of two parts, *eye* and *lash*, each of which relates to an independent word. The existence of parallel formations in the language (e.g. *eye-brow*, *finger-nail*, etc.) confirms the idea of the *regularity* of the relationship between *eye* and *lash*. Decomposable terms in some languages can correspond to non-decomposable ones in others, e.g. Portuguese *cílio* 'eyelash'. Similarly, non-decomposable items in English such as *beard* or *moustache* correspond to composite structures in Tariana (Arawak,