

**Typology and Evolution of Cardinal Numeral-Noun
Constructions**

Vipas Pothipath

B.A., M.A., MSc by Research

A thesis submitted in fulfilment of requirements for the degree of

Doctor of Philosophy

to

Linguistics and English Language

School of Philosophy, Psychology and Language Sciences

University of Edinburgh

March 2008

© Copyright 2008

By

Vipas Pothipath

To...my parents in heaven, Taweesak and Pilas, who taught me to count things

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledge is made in the text.

Vipas Pothipath

Abstract

The principles of economy and distinctness in language seem to be key selection pressures for language evolution. Accordingly, to express the exact number of things, humans might be expected to use cardinal numeral-noun constructions (CNNCs) consisting of just two constituents, namely a noun (N) representing quantified things and a cardinal numeral (NUM) representing the number of the quantified things (for example, English *three sheep*). However, the structural patterns of CNNCs used in a number of languages spoken today are not that simple, and have seemingly redundant constituents, typically non-singular markers (NSG) and numeral classifiers (CLF). CNNCs observed in the world's languages also appear to show a diversity of structural patterns despite the fact that simple structures like English *three sheep* seem very practical. This observation brings up two related major goals of this thesis. The first is to reveal structural types of cardinal numeral-noun constructions of singularity (CNNC_{SG}) and cardinal numeral-noun constructions of non-singularity (CNNC_{NSG}). The other major goal is to hypothesize a possible evolutionary scenario for CNNCs since their emergence till the modern era.

This thesis approaches these two issues by exploring CNNCs in 241 languages representing 101 language groups (i.e. language families, language isolates and pidgins and creoles) across the globe through reference grammars to ensure the greatest range of possible attested structural patterns of CNNCs. This cross-linguistic survey demonstrates that, with regard to CNNC_{SG}, the world's languages are divided into two major types, namely {N,NUM} and {N,NUM,CLF} with relatively few other possibilities. In relation to CNNC_{NSG}, the world's languages are divided into four major types, namely {N,NUM}, {N,NUM,NSG}, {N,NUM,CLF} and a mixture of {N,NUM} and {N,NUM,NSG} with some other less common possibilities. The historical origins of these structural types are then investigated, using evidence from old written records together with theoretical approaches, especially grammaticalization. Finally, it is found that the various structural patterns of CNNCs discovered can be considered in the light of a hypothetical evolutionary ladder. Hence, with cross-linguistic comparison integrated

with diachronic approaches, hypothesized evolutionary trajectories of CNNCs are postulated. It is conjectured that the construction consisting of a noun plus a word with a numerical interpretation such as the words meaning 'alone' or 'company' may represent a possible initial stage of CNNCs. From that stage onwards, CNNCs have split into many types over time. The development is reversible in terms of structural complexity, and idiosyncratic in some cases. Besides, the contributory factors in the development of CNNCs involve a quantifying function, a non-quantifying function, and a mixture of both.

Based on the study of evolution of CNNCs, this thesis also discusses the nature of language by comparing language change with biological evolution in some major aspects. The comparison suggests that language is strikingly similar to biological organisms in general, perhaps rather than to other cultural artefacts. Overall, this thesis contributes to current studies of the complexity and diversity of human language(s).

Acknowledgements

I wish to express my highest gratitude to my principal supervisor, Professor Jim Hurford for his impressive caring guidance and invaluable advice, above all, for his optimism and understanding throughout. I would also like to extend my gratitude to my second supervisor, Professor April McMahon, for her interest in my works and her time for helpful discussions and feedback despite of her busy schedule. My special appreciation goes to my examiners, Professor Greville Corbett and Dr. Andrew Smith for their constructive comments and detailed suggestions which have improved this thesis a great deal.

I am deeply grateful to Professor Bernard Comrie, Director of the Department of Linguistics, the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany for the permission to use the library facilities freely, which were extremely valuable for my data collection. The language data used in the thesis also owe much to many people who kindly answered my questions. These include Shirin Abadikhah, Anvita Abbi, Paolo Acquaviva, Willem Adelaar, Alexandra Aikhenvald, Giorgos Argyropoulos, Ronald Asher, Peter Austin, Philip Davis, Mark Donohue, Matthew Dryer, Mark Elliott, Nick Evans, Harry Feldman, William Gillies, Raquel Guirardello-Damian, Bernd Heine, George Hewitt, Luise Hercus, Clive Holes, Hae-Sung Jeon, Yuri Kato, Chonghyuck Kim, Marian Klamer, Thomas Lehmann, Augusto Lorenzino, Judith Maxwell, Miriam Meyerhoff, Merilin Miljan, Marianne Mithun, Dmitri Morenkov, Michael Noonan, Rachel Nordlinger, Knut Olawsky, Kyoko Otsuki, Peter Patrick, Gareth Roberts, Alan Rumsey, Gunter Senft, Paul Sidwell, Carsten Stuber, Brian Tawney, Victor Tron, Marit Vamarasi, Viveka Velupillai, Virve-Anneli Vihman, Craig Volker and Alexander Vovin.

Cordial thanks to Kate Naomi Juda for her inspiring question leading to the thesis topic; Jennifer Sullivan, Catherine Dickie and John Stevenson for their careful proofreading and very helpful suggestions; Phaisit Boriboon for very handy electronic books; Morag Brown and Michael Bennett for computing support; and my sister Wipasinee and my brother Khantichai for their various assistance throughout.

I am particularly indebted to all members of the Department of Thai, Faculty of Arts, Chulalongkorn University, Thailand for granting me sabbatical leave to pursue my studies at Edinburgh and for their generosity of dealing with all my workload during my five postgraduate years. In particular, I am grateful beyond words to Ajarn Suchitra Chongsatitwatana for her tremendous support and unwavering kind concerns.

My studies at Edinburgh would not have been possible without the support of the Royal Thai Government Scholarship, for which grateful acknowledgement is hereby made.

Abbreviations and Symbols

A	agent-like argument
ABS	absolute
ACC	accusative
ADJ	adjective
ADJZ	adjectivalizer
ADV	adverb
ADVZ	adverbializer
ANIM	animate
AOR	aorist
APPL	applicative
ART	article
ASP	aspect
ATTR	attributive
AUG	augmented
AUX	auxiliary
BAS	basic number
CL	class (noun class, gender)
CLE	locative copula (existential)
CLF	numeral classifier (including numeral classifier-like word)
CM	Comparative Method
CNNC _{NSG}	cardinal numeral-noun construction of non-singularity
CNNC _{SG}	cardinal numeral-noun construction of singularity
COLL	collective
COMP	complementizer
COP	copula
CP	creoles and pidgins
CPL	completive
DAT	dative

DECL	declarative
DEF	definite
DEICLF	deictic classifier
DET	determiner
DEX	indexer
DIM	diminutive
DPM	double plural marker
DST	distal
DU	dual
EMPH	emphatic
ERG	ergative
EXP	experiencer
EXT	extent of action
F	feminine
FOC	focus
GEN	genitive
HUM	human
IEXC	1 st person non-singular inclusive (including addressee)
IMP	imperative
IND	indicative
INDEF	indefinite
INS	instrumental
INV	inverse
IR	Internal Reconstruction
ITER	iterative
L	linker
LAT	lative
LOC	locative
LOCUT	locator person
LQ	lexical quantifier
M	masculine

MIN	minimal
MPA	point of action marker
MW	measure word
N	noun
NARR	narrative
NC	noun class
NM	non-masculine
NNI	noun number indicator
NNQEs	non-numeral quantificational expressions
NOM	nominative
NONHUM	non-human
NPL	non-plural
NPST	non-past
NSG	non-singular
NUM	numeral
NUMPCL	numerical particle
O	transitive object
OBL	oblique
PASS	passive
PAU	paucal
PCL	particle
PFV	perfective
PFX	preverbal prefix
PL	plural
PNUM	numeral marker for persons
POSS	possessive
POSSD	possessed
PRED	predicative
PREP	preposition
PRES	present
PREV	preverb

PROG	progressive
PRT	preterite (past)
PRTV	partitive
PST	past
QUOT	quotative
REDUP	reduplication
REL	relative
REP	repetition
RMS	subject relative clause marker
RP/P	realis past/present
SBJ	subject
SF	subject focus
SG	singular (or singulative)
SS	same subject
SUB	subordinator
SW	switch reference
T/A	tense/aspect
TAM	marker of tense, aspect, or modality
TNS	tense
TOD.PST	today's past tense
TOP	topic
TR	transitive
TRI	trial
VY	verbal stem
WALS	World Atlas of Language Structures
1	first person
2	second person
3	third person
I, II, III, etc.	Class I, Class II, Class III, etc.
-	morpheme boundary
=	clitic boundary

()	optional element, e.g. {N,NUM,(NSG)}; inherent category, e.g. French <i>homme</i> ‘man (M)’
.	portmanteau morpheme boundary
@	subsidiary (type)
[]	non-overt element (used in interlinear glosses, e.g. [NOM.SG])
[-]	resource unavailable
[+]	resource available
{ }	set of elements in a construction
↓	becomes
<	derives from
>	becomes

Contents

Declaration	iv
Abstract	v
Acknowledgements	vii
Abbreviations and Symbols	ix
1 Introduction	1
1.1 The fundamental dualism and cardinal numeral-noun constructions	1
1.2 Aims	4
1.3 Overview	6
1.4 Transcription and glossing	8
2 Cardinal Numeral-Noun Constructions and Earlier Research	10
2.1 Cardinal numeral-noun constructions: an illustration	10
2.2 Earlier research on typology and evolution of CNNCs	16
2.2.1 Sanches and Slobin's 'Numeral classifiers and plural marking'	16
2.2.2 Rijkhoff's 'Seinsarten'	18
2.2.3 Hurford's 'The interaction between numerals and nouns'	21
2.2.4 Conclusion	24
3 How to Trace the Evolution of CNNCs	25
3.1 Exploring types of CNNCs	25
3.1.1 Sampling	25
3.1.2 Data sources	27
3.2 Investigating historical origins of CNNCs	32
3.2.1 Evidence from old texts	32
3.2.2 Theoretical diachronic approaches	33
3.2.2.1 Grammaticalization theory	33
3.2.2.2 Lexicalization	37
3.2.2.3 Linguistic reconstruction	40
3.2.2.4 Language contact	47
3.3 Postulating the evolutionary trajectories of CNNCs	50
3.4 Conclusion	56
4 Before the Rise of Cardinal Numeral-Noun Constructions	57

4.1 Reduplication	59
4.2 Non-numeral quantifiers	64
4.3 Grammatical means of number distinction	65
4.4 Words with a numerical interpretation	67
4.5 The rise of numerals and the polygenesis of CNNCs	72
4.6 Conclusion	74
5 On the Extra Elements	75
5.1 Number markers	76
5.1.1 Number-marking systems	76
5.1.2 Singular and singulative	78
5.1.3 Plural	81
5.1.4 Other number markers	89
5.1.4.1 Dual marker	89
5.1.4.2 Trial marker	90
5.1.4.3 Non-plural marker	91
5.1.4.4 Inverse number marker	91
5.2 Numeral classifiers	93
5.2.1 Characteristics of numeral classifiers	93
5.2.2 Numeral classifier languages	95
5.2.3 Lexico-grammatical items in the numeral classifier slot	96
5.2.3.1 Measure words	96
5.2.3.2 Noun classifiers	97
5.2.3.3 Noun class (or gender) markers	99
5.2.3.4 Number markers	100
5.2.4 A frequently cited view on the genesis of numeral classifiers	101
5.3 Other quantificational extra elements	104
5.3.1 Relative clause marker	104
5.3.2 Numerical particle	105
5.3.3 Accusative case marker	107
5.3.4 Oblique marker	107
5.3.5 Numeral marker for persons	110
5.3.6 Double plural Marker	110
5.4 Non-quantificational extra elements	111
5.4.1 Noun class (gender)	111
5.4.2 Copula	111

5.4.3	Attributive	112
5.4.4	Switch reference	113
5.5.	Conclusion	114
6	CNNCs Across Languages	115
6.1	Types of CNNC _{SG}	117
6.2	Geographical distribution of CNNC _{SG}	125
6.3	Types of CNNC _{NSG}	129
6.4	Geographical distribution of CNNC _{NSG}	142
6.5	Conclusion	147
7	CNNCs and Noun Class: An Implicational Universal?	148
7.1	What motivates the hypotheses?	150
7.2	How can we test an implicational universal?	153
7.3	Testing the hypothesised implicational universals	160
7.4	Why is {N,NUM,NSG,(X)} associated with noun class?	166
7.5	Conclusion	169
8	Historical Origins of CNNC _{SG}	170
8.1	{N,NUM}	172
8.1.1	{N,NUM} < N + 'one-ness'	172
8.1.2	{N,NUM} < {N,NUM,CLF}	172
8.2	{N,NUM,SG}	179
8.2.1	{N,NUM,SG} < {N,SG} < {N,NUM}	179
8.2.1.1	Burushaski	179
8.2.1.2	Pame	181
8.2.2	{N,NUM,SG} < {N,SG} < N+third person singular pronoun	184
8.3	{N,NUM,CLF}	187
8.3.1	{N,NUM,CLF} < {N,NUM}	187
8.3.1.1	Chinese	187
8.3.2	{N,NUM,CLF} < {N,NUM,CLF,SG}	198
8.4	{N,NUM,CLF,SG}	202
9	Historical Origins of CNNC _{NSG}	205
9.1	{N,NUM}	205
9.1.1	{N,NUM} < N + 'two-ness'	205
9.1.2	{N,NUM} < {N,NUM,NSG}	206
9.1.2.1	Celtic (Irish and Welsh)	206

9.1.2.2 Creoles and Pidgins	207
9.1.3 {N,NUM} < {N,NUM,CLF}	210
9.2 {N,NUM,NSG}	214
9.2.1 {N,NUM,NSG} < {N,NSG} where NSG is plural	214
9.2.2 {N,NUM,NSG} < {N,NSG} where NSG is dual	216
9.2.2.1 Arabic	216
9.2.2.2 Futuna-Aniwa	218
9.2.3 {N,NUM,NSG} < {N,NUM}	219
9.2.4 {N,NUM,NSG} < {N,NUM,OBL,NSG}	220
9.2.4.1 English	220
9.2.4.2 Arabic	222
9.3 {N,NUM,CLF}	224
{N,NUM,CLF} < {N,NUM,NSG}	224
9.4 {N,NUM,CLF,NSG}	225
9.4.1 {N,NUM,CLF,NSG} < {N,NUM,NSG}	225
9.4.2 {N,NUM,CLF,NSG} < {N,NUM}	227
9.5 {N,NUM,OBL,SG}	231
9.5.1 {N,NUM,OBL,SG} < {N,NUM,NSG}	231
9.5.2 {N,NUM,OBL,SG} < {N,NUM}	232
9.6 {N,NUM,OBL,NSG}	236
9.6.1 {N,NUM,OBL,NSG} < {N,NSG}	236
9.6.2 {N,NUM,OBL,NSG} < {N,NUM}	239
10 Evolutionary Trajectories of CNNCs	241
10.1 Evolutionary trajectory of CNNC _{SG}	242
10.2 Evolutionary trajectories of CNNC _{NSG}	246
10.3 Motivations for the origins of CNNCs	251
10.3.1 Quantifying function-motivated change	251
10.3.2 Non-quantifying function-motivated change	251
10.3.3 Quantifying function-motivated change plus non-quantifying function-motivated change	257
10.3.4 Summary	258
10.4 How did CNNCs evolve?	259
10.5 Conclusion	261
11 The True Nature of Language: A Perspective from the Evolution of CNNCs	262

11.1 Analogy: What and Why? _____	262
11.2 Misuses and sensible uses of biological analogies _____	267
11.2.1 Misuses of biological analogies _____	267
11.2.2 Sensible uses of biological analogies _____	272
11.3 Analogies between organisms and linguistic units _____	278
11.3.1 Evolution: descent with modification _____	278
11.3.2 Origins _____	279
11.3.3 Evolutionary trend _____	280
11.3.4 Gradual change and sudden change _____	282
11.3.5 Speciation _____	284
11.3.6 Mechanisms of evolution _____	286
11.3.7 Adaptation, vestigial structures, maladaptation, and exaptation _____	297
11.3.8 Convergent evolution _____	297
11.3.9 Uniformitarianism _____	298
11.4 Conclusion _____	302
12 Conclusions _____	304
12.1 Summary of findings _____	305
12.1.1 Typological issues of CNNCs _____	305
12.1.2 Historical origins of CNNCs _____	307
12.1.3 Evolutionary scenario for CNNCs: a possibility _____	308
12.1.4 Explanations for the complexity and diversity of CNNCs _____	310
12.2 Concluding remarks _____	313
Appendix 1 Lists of Languages (by area) and Structural Types of CNNCs _____	316
Table A: Sample languages and their structural types of CNNC _{SG} _____	317
Table B: Sample languages and their structural types of CNNC _{NSG} _____	326
Appendix 2 Lists of Languages _____	338
2.1 List of sample languages _____	338
2.2 Lists of languages for Table 6.1 _____	341
2.3 Lists of languages for Table 6.2 _____	344
2.4 Lists of languages for Table 7.1 _____	348
2.5 Lists of languages for Table 7.9-Table 7.13 _____	350
Appendix 3 Survey of Reference Grammars and Data _____	358
References _____	617

Index of Tables

Table 5.1 Verbal prefixes for numerals in Oceanic languages _____	105
Table 6.1 Language types with regard to $CNNC_{SG}$ and their frequencies _____	117
Table 6.2 Language types with regard to $CNNC_{NSG}$ and their frequencies _____	129
Table 7.1 The relationship between plural and gender _____	151
Table 7.2 Property Q in languages with property P _____	154
Table 7.3 Property Q in languages with property -P _____	155
Table 7.4 Proportions of genera with property Q among languages with P and -P ____	156
Table 7.5 Property Q in languages with property P _____	156
Table 7.6 Property Q in languages with property P _____	157
Table 7.7 Property Q and Property P _____	158
Table 7.8 Proportions of genera that are Q among languages with P and -P ____	159
Table 7.9 Property [+NC] in languages with + {N,NUM,NSG,(X)} _____	160
Table 7.10 Property [+NC] in languages with -{N,NUM,NSG,(X)} _____	161
Table 7.11 Proportion of genera which have the property [+NC] _____	162
Table 7.12 Languages with -{N,NUM,NSG,(X)} and the property [+NC] _____	163
Table 7.13 Languages with +{N,NUM,NSG,(X)} and the property [-NC] _____	164
Table 7.14 Proportions of genera which have the property [-NC] _____	165
Table 8.1 Six classes of numeral 'one' in Teribe _____	177
Table 9.1 Two sets of numerals in Nivkh _____	213

Index of Figures

Figure 3.1 Morphological reconstruction of the lexical form ‘person’ in Jukun	45
Figure 3.2 Basis of reconstruction	50
Figure 3.3 Basis of reconstruction: co-existent types	51
Figure 3.4 Basis of reconstruction: two types derived from the same source	51
Figure 3.5 Basis of reconstruction: same types derived from different sources	52
Figure 3.6 Basis of reconstruction: two different origins of the current types	52
Figure 3.7 Layers of grammatical development	54
Figure 4.1 Hierarchy of semantic properties expressed by reduplication	62
Figure 5.1 Degrees of obligatoriness of number marking	82
Figure 5.2 Representational structure of counting collective references	109
Figure 8.1 Process of change in $CNNC_{SG}$ in Burushaski	180
Figure 8.2 Process of change in $CNNC_{SG}$ in Pame	182
Figure 8.3 Process of change in $CNNC_{SG}$ in Lunda	185
Figure 9.1 Continuum of compactness of numeral plus numeral classifier	211
Figure 10.1 A Possible Evolutionary Trajectory of $CNNC_{SG}$	242
Figure 10.2 A Possible Evolutionary Trajectory of $CNNC_{NSG}$	246
Figure 11.1 Schema of analogy between the solar system and the atom	265
Figure 11.2 Phylogenetic schema: German, English and French	271
Figure 11.3 Implicational hierarchy of basic colour terms	273
Figure 11.4 The skulls of the Beluga Whale and its ancestors in different periods	282
Figure 11.5 Speciation events in the partial <i>Drosophila</i> phylogeny	284
Figure 11.6 Splitting event of a partial phylogeny of Indo-European $CNNC_{NSG}$	285
Figure 12.1 A Possible Evolutionary Trajectory of $CNNC_{SG}$	308
Figure 12.2 A Possible Evolutionary Trajectory of $CNNC_{NSG}$	309

Index of Maps

Map 4.1 Geographical distribution of reduplication	61
Map 6.1 Geographical distribution of CNNC _{SG}	125
Map 6.2 Western Africa (zoomed)	125
Map 6.3 New Guinea and Northern Australia (zoomed)	125
Map 6.4 Geographical distribution of CNNC _{NSG}	142
Map 6.5 New Guinea and Northern Australia (zoomed)	142
Map 6.6 Mesoamerica (zoomed)	142
Map 6.7 Geographical distribution of {N,NUM,CLF} with respect to degrees of obligatoriness	143
Map 9.1 Geographical distribution of numeral classifiers in Sino-Tibetan	229

1 Introduction

Linguistic development follows not one tendency, but two opposing ones: towards distinctness and towards economy. [...] Two opposed interests wrestle with each other, are both active and effective at the same time and alternate in predominance [...] (Leopold 1930: 102-104)

1.1 The fundamental dualism and cardinal numeral-noun constructions

There has long been a widespread view in the general linguistics literature that the *principle of economy* (alias *simplicity*) in language is a significant characteristic of human communication. This is mentioned in several places with various well-known versions, such as Zipf's (1949/1965) *principle of least effort*, Grice's (1975) *maxim of quantity*, and Dahl's (2004) *principle of redundancy management*. According to several versions, the term *principle of economy* can be defined as the principle that speakers use the least amount of energy or effort in speech to achieve satisfactory results. A simple example of the economy principle can be observed in the minimization of articulatory effort. For example, the English word *exam* in spoken English is shortened from the original word *examination*. In this case, the last three syllables (i.e. *-ination*) are clipped because the speaker tries to spend the least amount of energy in articulation by shortening the word when the unclipped part is found sufficient to indicate the whole. As the economy principle is a significant characteristic of human communication, the principle is frequently claimed to be a primary force contributing to diachronic change in language (such as the birth of the word *exam* in English) or language evolution as a whole.

However, ideally the economized message that the speaker sends must be unambiguous or not too vague for the hearer. Otherwise the communication may not be successful. For that reason, human language also tends to evolve towards another principle generally known as *distinctness* (alias *expressivity* or *clarity*) when the speaker finds that the language structure used in a given time is not sufficient. To illustrate, in certain languages, the noun is vague in terms of number. For example, in Yoruba, a Niger-Congo language spoken in Benin and Nigeria, the noun *ajá* ‘dog’ can refer to one dog or more than one dog. In the situation where the number of dogs is important or needs emphasis, the speaker then makes the form distinctive by attaching the pronoun *áwọn* ‘they’ to the noun (Rowlands 1993: 40). The use of linguistic elements to distinguish number in nouns such as this may make the morpho-syntactic system in the language become more complex. Accordingly, language complexity is attributed to the principle of distinctness.

Together, the economy and distinctness principles are referred to in Leopold (1930: 1) as the “fundamental dualism” towards which human language tends to develop over time. These two principles in fact lie in the unified principle of *communicative efficiency*, which assumes that the speakers tend to “[m]aximize the chances of safe delivery and minimize costs [of the message]” (Dahl 2004: 11). When the language system becomes superfluous (i.e. communication is costly), speakers will manage to economize on it. Once the language system becomes so economical that it is not clear enough, the speakers will find the compensatory devices to clarify it. Hence, human language is generally assumed to evolve towards the balance of economy and distinctness.

However, in actual practice, a number of cases show that human language does not always maintain the balance between the two principles. One such counterexample can be observed in quantificational expressions—expressions of the number of things. The syntactic constructions expressing the *exact* number of things are referred to in this thesis as *cardinal numeral-noun constructions* (CNNCs). Following the two principles of economy and distinctness, humans would be expected to use quantificational

expressions consisting of just two linguistic elements, namely a noun denoting things and a cardinal numeral denoting numbers as in (1.1).

- (1.1) *Kpelle (Niger-Congo; Liberia; Welmers 1973: 294)*
- | | |
|---------------------|--------------|
| <i>pére</i> | <i>feere</i> |
| <i>house</i> | <i>two</i> |
| <i>'two houses'</i> | |

CNNCs comprising just two constituents, namely a cardinal numeral (e.g. *two*) and a noun, as in *Kpelle*, are referred to as *Simple CNNCs*. However, in a large number of languages, the CNNCs are more complex due to seemingly redundant elements, typically a *non-singular marker* (NSG) (i.e. a linguistic form, most often an affix, indicating number above *one*, typically of nouns, e.g. the English suffix *-s* in *two dogs*) and a *numeral classifier* (CLF) (i.e. a linguistic form normally appearing next to a numeral or a non-numeral quantifier, such as *many*, categorizing the noun with which it co-occurs on a semantic basis) as in Thai in (1.2). These types of CNNCs are referred to as *Complex CNNCs* (more detailed descriptions of CNNCs and the seemingly redundant elements are given in §2.1 and Chapter 5 respectively).

- (1.2) *Thai (own knowledge)*
- | | | |
|---------------------|--------------|--------------------------|
| <i>mǎ:</i> | <i>sa:m</i> | <i>tua</i> |
| <i>dog</i> | <i>three</i> | <i>CLF (lit. 'body')</i> |
| <i>'three dogs'</i> | | |

This general observation seems to violate the principle of economy in language evolution and in addition, the principle of distinctness does not seem to account for the complexity. This observation leads to the stimulating questions of 1) how the other types of CNNCs came into existence despite the fact that Simple CNNCs are capable of expressing numerical quantification, and 2) how the two principles will explain this situation.

1.2 Aims

The question above brings up a set of goals for this thesis. The first is to reveal the types, distribution and frequencies of CNNCs throughout the world's languages. We will then find out whether or not Simple CNNCs are the most common compared to other categories of CNNCs. The findings of the cross-linguistic study are also necessary for investigating the evolutionary paths of CNNCs because the investigation requires knowing what possible types of CNNCs may be used. What becomes particularly interesting in the light of cross-linguistic studies or *typology* is whether or not we can correlate the types of CNNCs with other linguistic features. Also, if a correlation does exist, an aim is to investigate further whether it can be attributed to typological, genetic or areal phenomena.

In addition, so far the scenario for the history of CNNCs in languages has not been fully described; the project therefore ultimately aims to outline the history of CNNCs since their emergence right up until the modern era, with the emphasis on a possible evolutionary scenario for CNNCs as well as the contributory factors driving the developments. This part will contribute to the understanding of the reasons why a number of languages also choose Complex CNNCs in addition to Simple CNNCs, and to what extent the variety of CNNCs is entirely associated with the economy and distinctness principles. Finally, the study of language change is generally regarded as an important way to understand the nature of language, and therefore the study of the evolution of CNNCs may, we hope, aid our understanding of certain aspects of the nature of language.

As suggested by the term *cardinal* in *cardinal numeral-noun constructions*, a limitation in scope of this thesis is the exclusion of other related constructions containing other kinds of numerals. There are many types of numerals, for example, *ordinal numerals* (i.e. numerals showing a position in a set of numbers, e.g. English *first, second, third*), *approximative numerals* (i.e. numerals which are close to the exact number, perhaps a little more or less, e.g. English *five or six, thousands*), *distributive*

numerals (i.e. numerals showing each member of the group, e.g. English *one each*), and so on. These numerals when combined with nouns may require some other linguistic elements and these constructions have their own meanings. For example, the construction containing approximative numerals plus nouns in Quechua (Huallaga) (Quechuan; Peru) consists of three elements, namely a noun and two numerals as in (1.3).

- (1.3) *Quechua (Weber 1989: 253)*
- | | | |
|---------------|--------------|--------------|
| <i>ishkay</i> | <i>kimsa</i> | <i>hunaq</i> |
| <i>two</i> | <i>three</i> | <i>day</i> |
- 'two or three days.'*

Example (1.3) is not treated as a CNNC because the meaning of the construction is not the same as that of a CNNC. The constructional meaning of CNNCs concerns the *exact* number of the referent, whereas the number of days expressed in (1.3) is either *two* or *three*. Like CNNCs, these constructions may have a variety of structural patterns also and they can be intensively scrutinized from typological and diachronic perspectives as well. However, such an investigation should be another piece of research work. Otherwise, it would have required that the scope of this thesis be greatly expanded and beyond the time limit of the project. Therefore, the constructions containing these other kinds of numerals will not be studied here.

1.3 Overview

The thesis consists of four major parts. In Part I, the first four chapters set out some preliminary issues as a necessary background for understanding the current project. After this introductory chapter, Chapter 2 provides the definition of CNNCs as well as a detailed description of the categories of CNNCs. In this chapter, previous studies relating to CNNCs are also reviewed. Then, Chapter 3 describes a procedure for investigating the evolution of CNNCs. Since the procedure for tracing the evolutionary paths of CNNCs is similar to the *Intergenetic Grammaticalization Comparison* approach employed in Heine and Kuteva (2002b, 2007), the said approach is also discussed in this chapter. In addition, this chapter introduces the theoretical diachronic approaches relevant to the current research, namely grammaticalization, lexicalization, linguistic reconstruction and language contact. Chapter 4 presents a speculative scenario for quantificational expressions before the emergence of CNNCs. This chapter describes various types of non-numeral quantificational expressions (NNQEs) in modern human languages. The NNQEs are conjectured to be the stage before the rise of CNNCs in the history of quantificational expressions.

Part II, which is a core part of the thesis, presents a study of CNNCs in a typological perspective. This part consists of three chapters. Treated as the introductory chapter to Part II, Chapter 5 provides the working definitions of the non-core elements of CNNCs such as number markers and numeral classifiers. Chapter 6 illustrates various language types of CNNCs along with their global distribution. After the description of language types and their distribution, Chapter 7 moves further to another typological task, namely *typological generalizations*. This chapter examines the correlation between the language types of CNNCs and noun class systems. An implicational universal is then proposed. The implicational universal proposed is tested with a test for the statistical significance of implicational universals designed by Dryer (2003). The testing method is also reviewed in this chapter.

Part III is another core part of the thesis, containing three chapters. This Part presents a study of CNNCs in a diachronic perspective. In Chapter 8 and Chapter 9, the structural patterns of CNNCs which have been reported in Chapter 6 are further investigated to examine their historical origins as well as contributory factors. Chapter 8 deals with the historical origins of structural patterns of *cardinal numeral-noun constructions of singularity* (e.g. English *one dog*) and Chapter 9 deals with the historical origins of *cardinal numeral-noun constructions of non-singularity* (e.g. English *two dog-s*), along with the contributory factors in the origins of those structural patterns. Next, Chapter 10 combines together all the various historical pathways of CNNCs to postulate a possible general evolutionary trajectory for CNNCs. Also, this chapter looks into the contributory factors again in terms of quantifying functions to see how and why CNNCs evolved. The last section discusses two aspects of the evolution of CNNCs, namely *directionality* (i.e. whether the diachronic developments are directional or reversible) and *regularity* (i.e. whether the diachronic developments are regular or idiosyncratic).

Finally, Part IV consists of two chapters. Chapter 11 is a reflection. It attempts, on the basis of having seen the evolution of CNNCs, to answer the question of what we have learned about the *true nature* of language from the study of evolution of CNNCs. Assuming that *nature* in language change can be understood by analogy with (post-) Darwinian evolutionary concepts, this chapter thus compares the two evolutionary systems to see the similarities and differences between language and biological organisms, using evidence from the study on CNNCs. Finally, Chapter 12 summarizes the thesis and adds a few concluding remarks.

1.4 Transcription and glossing

Almost all the examples provided in this thesis are taken from reference grammars. The authors of these grammars use different styles of transcription and glossing. Regarding the transcription, some use a Romanised transcription and others use a phonetic transcription. The authors' preferences are respected as much as possible, so the transcriptions of the examples are copied without change from the original texts. As far as glossing is concerned, although the glosses of the examples used mostly follow those of the original source, a few changes to some examples are made to unify the glossing systems in the thesis. The changes involve the uses of abbreviated grammatical category labels and symbols as follows.

Firstly, when different sources use different abbreviated labels for the same grammatical category (e.g. in some sources, the dual marker is represented by DL while others represent it with DU), the one which conforms to the standard abbreviations offered in *The Leipzig Glossing Rules* (Comrie, Haspelmath and Bickel 2008) and/or to the conventions commonly used in linguistics literature is chosen.

Secondly, when the abbreviated grammatical category labels used by different authors appear to share the same form, but represent different concepts, new abbreviated labels are made for the rarer grammatical category (which is not present in the standard source such as *The Leipzig Glossing Rules*) to avoid ambiguity. For example, the abbreviation *NUM* in most reference grammars represents *numeral* but in a few reference grammars, such as Maori (cf. Bauer, Parker and Evans 1993), the abbreviation *NUM* represents *numerical particle*. In this case, the newly created label (i.e. NUMPCL) is used instead.

Thirdly, in some examples, the authors use words from the metalanguage like English for grammatical morphemes, such as glossing the definite article by using the English article *his* instead of the abbreviated label *3SG.POSS*. Although either an abbreviated grammatical category label or a word from metalanguage is acceptable in many cases (cf. Comrie, Haspelmath and Bickel 2008: 3), for a consistency reason, the thesis prefers to use only one system, namely abbreviated grammatical category labels.

In these cases, those English glosses for grammatical morphemes are changed into corresponding grammatical category labels. For practical reasons, however, this principle excludes the metalanguage for prepositions, such as *of*, *from*, *with*, and *for*, which are kept as used in original sources.

Fourthly, in relation to the non-overt elements such as nominative case and inherent categories such as gender/noun class, if known from the source, following *The Leipzig Glossing Rules*, the non-overt elements and inherent categories are enclosed in square brackets and the round parentheses respectively. However, in the case that the non-overt elements and inherent categories are not clearly indicated in the source, they will be treated as a portmanteau morpheme (i.e. a morpheme which contains multiple grammatical and/or semantic elements) and will be separated by a period in the gloss.

Fifthly, some numeral classifier-like words (e.g. English *head* in *two hundred head of cattle*, cf. §5.1.2) are glossed with the grammatical category label *CLF* along with its literal meaning (e.g. *CLF 'head'*), although they are not numeral classifiers proper.

In addition, the interlinear glosses for plural (PL), dual (DU), trial (TRI) are generalised as *non-singular* (NSG) in set notations. Therefore, the noun suffixed with a dual marker (i.e. N-DU) will be encoded as *{N,NSG}* in set notation when referring to structural patterns of CNNCs, unless otherwise specified. Also, the interlinear glosses for prepositions, partitive case (PRTV), and genitive case (GEN) are generalised as *OBL* in set notations when referring to structural patterns of CNNCs.

Moreover, in relation to gender/noun class glossing, different authors again use different styles. In this case, the thesis uses the Roman numeral to indicate the class number (if known from the source). If the gender/noun class number is not clear from the source, the thesis uses the abbreviated category label *CL* to signify that the morpheme is gender/noun class.

Finally, different sources use different symbols (e.g. colon (:), plus (+), period (.) to separate multiple grammatical and semantic elements in one word. In this case, the thesis prefers to collapse all those symbols in one, namely a period (.)

2 Cardinal Numeral-Noun Constructions and Earlier Research

2.1 Cardinal numeral-noun constructions: an illustration

Cardinal numeral-noun constructions (CNNCs) are syntactic constructions basically consisting of two core constituents, namely a cardinal numeral X and a noun Y , mostly without regard for the order in which they occur.¹ Cardinal numerals are the words which show the number of entities, such as English *one* in *one man*. The meaning of the construction is *the number X of Y* . In the example of English *one man*, the word *one* is the cardinal numeral and the word *man* is the noun, and the meaning of the construction is a set of individuals of the type denoted by *man*, and of numerosity 1.

CNNCs can be classified into two categories based on the number of the noun Y referred to. The first category is *cardinal numeral-noun constructions of singularity* (CNNC_{SG}), referring to the cardinal numeral-noun constructions in which the number of the noun Y is *one*. The second category is *cardinal numeral-noun constructions of non-singularity* (CNNC_{NSG}) referring to cardinal numeral-noun constructions in which the number of the noun Y is greater than *one*. CNNC_{SG} are realized with a cardinal numeral equivalent to *one*, while CNNC_{NSG} are realized with the cardinal numeral greater than *one* as exemplified by Kpelle (Niger-Congo; Liberia) in (2.1a) and (2.1b) respectively.

¹ However, for the languages in which the same numerals can be both cardinals and ordinals, the order of the numerals and nouns may constitute different constructions. Examples are provided by Sapuan (Austro-Asiatic; Jacq and Sidwell 1999: 30-31 in Stolz and Veselinova 2005: 218): (a) represents a cardinal numeral-noun construction, whereas (b) represents an ordinal numeral-noun construction.

- a. *səm* *bar* *lāj*
 house two CLF (for 'house')
 'two houses'
- b. *bar* *māj* *law*
 two language Lao
 'my second language is Lao'

(2.1) *Kpelle* (Welmers 1973: 294)

a. *péɛ* *tɔ̃nɔ*
 house *one*
 ‘one house’

b. *péɛ* *feere*
 house *two*
 ‘two houses’

In addition, based on structural complexity, CNNCs can be classified into three categories, namely Simple CNNCs, Complex CNNCs and Simplex CNNCs.

Simple CNNCs comprise just two core constituents, namely a cardinal numeral and a noun, as in (2.1). They also include a construction in which a noun combines with the dual marker (DU) (i.e. a linguistic form, typically an affix, referring to two) or the trial marker (TRI) (i.e. a linguistic form, typically an affix, referring to three), without being accompanied by free numerals two or three respectively. Simple CNNCs of this kind are illustrated by *Kuot*, spoken in New Ireland, Papua New Guinea, and *Arabana*, an Australian language of South Australia.

(2.2) *Kuot* (Chung and Chung 1996:42)

dagar-βien
egg-DU
‘two eggs’

(2.3) *Arabana* (Hercus 1994: 64)

mathapurda-karikari
old.man-TRI
‘three old men’

In (2.2), where the number of the noun referred to is two, the $CNNC_{NSG}$ is formed by just two constituents, namely the noun *daga* ‘egg’ and the dual suffix *βien* without being accompanied by the free numeral *two*. In (2.3), where the number of the noun referred to is three, the $CNNC_{NSG}$ is also composed of just two constituents,

namely the noun *mathapurda* ‘old man’ and the trial suffix *karikari*; the free numeral *three* is not required. The examples from Kuot and Arabana are therefore instances of Simple CNNCs. The types of CNNCs as illustrated in (2.1)-(2.3) above all count as Simple CNNCs because apart from the core elements, they contain no other elements.

Complex CNNCs are CNNCs which contain a noun, a numeral, and a certain non-core constituent (henceforth *extra element*), such as a non-singular marker or a numeral classifier. The said extra elements can be morphologically segmented when attached to the core elements. Complex CNNCs with a non-singular marker (NSG) and a numeral classifier are illustrated by the examples in (2.4) and (2.5) from French and Thai respectively.

(2.4) *French (own knowledge)*
trois chien-s
three dog-PL
 ‘three dogs’

(2.5) *Thai (own knowledge)*
mǎ: sǎ:m tua
dog three CLF (lit. ‘body’)
 ‘three dogs’

In French, the form of nouns undergoes certain changes, typically being suffixed with *-s*, to show non-singularity when reference is being made to more than one entity. In (2.4), the CNNC_{NSG} in French consists of the noun *chien* ‘dog’, the numeral *trois* ‘three’, and the non-singular suffix *-s*. On the other hand, in Thai, in both CNNC_{SG} and CNNC_{NSG}, a numeral classifier is obligatorily required. Thus, the CNNC_{NSG} in Thai as shown in (2.5) consists of the noun *mǎ:* ‘dog’, the numeral *sǎ:m* ‘three’, and the numeral classifier *tua* ‘body’ (a numeral classifier for nouns denoting animals, clothing, furniture, etc.).

The extra elements such as non-singular markers and numeral classifiers seem surplus to requirements of CNNCs—in other words, without them the expressions

illustrated in (2.4) and (2.5) remain comprehensible. Other elements can also add to the structural complexity of CNNCs, some of which play certain roles in expressing quantification, whereas others do not. The various types of the extra elements in CNNCs are discussed in detail in Chapter 5.

The last category is *Simplex CNNCs*. Like Simple CNNCs, these are CNNCs consisting of two core constituents, namely a noun and a cardinal numeral. However, Simplex CNNCs are different from Simple CNNCs in that the numeral in Simplex CNNCs is varied according to the noun it modifies and vice versa. This is because there seems to be an extra element, namely a number marker or a numeral classifier, inherently existing within the noun or the numeral respectively. These extra elements cannot be morphologically segmented. Where the number marker seems to be amalgamated into the noun, as in English *two men* (instead of **two man-s*), the construction illustrated by the English *two men* counts as an example of Simplex CNNC_{NSG} due to the fact that the non-singular marker in the non-singular noun is semantically fused with the noun in one word as a portmanteau word.

In some languages with numeral classifiers, a numeral classifier may be fused with a numeral. It is noted in Aikhenvald (2000: 108) that this phenomenon often occurs in languages with *fusional* and *polysynthetic* characteristics (i.e. the words are built up by two or more morphemes and the boundaries between the morphemes are fuzzy). An example is shown from Nivkh (isolate; southeastern Siberia and Sakhalin Island).

- (2.6) *Nivkh* (Gruzdeva 1998: 62)
- | | |
|------------------|----------------|
| <i>n'ivyn</i> | <i>menɲ</i> |
| <i>man</i> | <i>two.CLF</i> |
| <i>'two men'</i> | |

In (2.6), the CNNC_{NSG} in Nivkh consists of two constituents, namely the noun *n'ivyn* 'man' and the numeral *menη* 'two'. However, the numeral *menη* 'two' also contains the numeral classifier for human nouns. The existence of the implicit numeral classifier in the numeral *menη* can be proved by comparing the numeral form *menη* with the numeral form *mor* 'two'. Both refer to 'two' but the former is used for counting nouns denoting humans, whereas the latter is used for counting nouns denoting animals (Gruzdeva 1998: 24). Therefore, there exists a meaningful unit functioning as the numeral classifier fused with the numeral. As a result of the amalgamation of the numeral and the numeral classifier in those languages, CNNCs seem deceptively simple, while the numeral system turns out to be complex as there is more than one set of cardinal numerals. For instance, in Nivkh, the numerals are divided into 26 classes with different numeral forms for each class as exemplified above (Gruzdeva 1998: 23).

In addition, examples like English *two men* and Nivkh *n'ivyn menη* (lit. man two.CLF) 'two men' are not treated as Complex CNNCs because they do not overtly contain the analyzable extra elements which are the defining characteristic of Complex CNNCs.

The categorization of CNNCs in the world's languages is yet to be described. In the survey of about 250 languages, which we carried out, an effort has been made to apply consistent criteria in categorizing CNNCs. Since we shall look at a broad range of languages to investigate the structural variation of CNNCs and their development with respect to complexity, the categorization of CNNCs as just briefly outlined is therefore made as a framework for data analysis as will be shown in the following chapters.

Various other terms have been used in the literature to refer to cardinal numeral-noun constructions, such as *numeral phrases* (e.g. Corbett 2000: 36), *numeral-noun constructions* (e.g. Hurford 2003: 567, Aikhenvald 2000: 116), *numerical expressions* (e.g. Greenberg 1978: 272) and *quantificational expressions* (e.g. Corver et al. 2007: 1). However, limitations with each of these terms make it preferable to use the term CNNCs instead.

For instance, the term *numeral phrases* cannot be applied to a couple of languages in the current sample in which the numeral functions as a predicate of the noun subject. Accordingly in these languages, CNNCs are a sentence, not a phrase. For example, in Haida (isolate; Canada) as illustrated in (2.7) below, the numeral *sdiŋ* ‘be two’ is the verb of the noun subject *qwaay* ‘rope’ and so can be inflected for tense.

- (2.7) *Haida* (Hori 2001: 144)
qwaay sGa-sdiŋ- gən
rope CLF-be.two-PST
‘There were two pieces of rope.’

As for the terms *numeral-noun constructions* and *numerical expressions*, these labels are too broad for the phenomenon under investigation in the current project. As already mentioned in §1.2, the current project deals only with cardinal numerals, but these two terms may include other kinds of numerals such as *ordinal numerals*, and the terms accordingly include other related constructions which are not in the scope of the thesis.

In addition, the term *numerical expressions* used by Greenberg (1978: 272) is used in the sense describing the way the numerals are combined. For example, the numeral *sixteen* in Italian is expressed as *sei-dici* [lit. six-ten] ‘sixteen’.

Finally, the term *quantificational expression* is used in the context of the broader system of quantification. It includes non-numeral quantifiers, such as English *few/several/many*. The term is also adopted in this thesis when referring to the broad system of quantification covering non-numeral quantificational expressions (e.g. ‘many’ plus noun; see Chapter 4 for further discussion) and numeral quantificational expressions (e.g. cardinals or ordinals plus noun).

In this thesis therefore, the term *cardinal numeral-noun constructions* is employed to make it clear that the quantificational constructions under study involve cardinal numerals only.

2.2 Earlier research on typology and evolution of CNNCs

The whole picture of the typology and evolution of CNNCs has not yet been drawn. Despite this, various structural types of CNNC_{NSG} have been recognized sporadically as illustrated below. Also, it seems that the typology and evolution of CNNC_{SG} has been largely ignored in the general linguistic literature. In this section, those few earlier research works dealing with typology and evolution of CNNC_{NSG} are reviewed chronologically.

2.2.1 Sanches and Slobin's 'Numeral classifiers and plural marking'

A central observation relating to the typology and evolution of CNNCs probably first appeared in Sanches and Slobin's (1973) *Numeral Classifiers and Plural Marking: An Implicational Universal*. The paper proposes the hypothesis about the complementary distribution of the obligatory plural markers and numeral classifiers in languages as shown below.

If a language includes numeral classifiers in its dominant mode of forming quantification expressions, then it will also have facultative expression of the plural. In other words, it will not have obligatory marking of the plural on nouns (Sanches and Slobin 1973: 4, emphasis original).

This means that obligatory plural markers and numeral classifiers are not supposed to co-exist in a given language. To support the hypothesis, Sanches and Slobin explored a sample of some 70 languages across the world. The finding appears to suggest that the hypothesis is justified, as languages in which the two systems co-exist are extremely rare. In addition, the paper has certain arguments relevant to the study of typology and evolution of CNNCs.

In terms of typology, languages are divided into 4 types with regard to the presence of plural markers (PI) and numeral classifiers (CLF). They are [+PL/+CLF],

[+PL/-CLF], [-PL/+CLF] and [-PL/-CLF], where the symbol (+) means that the feature is available; and the symbol (-) means that the feature is not available. Hence, the notation [+PL/-CLF] means that the language has plural markers but *not* numeral classifiers, for instance. Note that the four types, in fact, do not really correspond to the types of CNNCs. That is, languages with plural markers do not necessarily use plural markers in CNNCs. Hungarian (Uralic; Hungary), for example, has plural markers but the plural markers are not used in CNNCs (Kenesei, Vago and Fenyvesi 1998: 39). So, the types [+PL/+CLF] and [+PL/-CLF] are not necessarily equivalent to {N,NUM,NSG,CLF} and {N,NUM,NSG} respectively.

In terms of evolution, it is argued that languages belonging to [-PL/-CLF] can develop or borrow *either* plurals (i.e. changing into [+PL/-CLF]) *or* numeral classifiers (i.e. changing into [-PL/+CLF]) rather than the other way round. To generalize, languages tend to undergo changes from simplicity (i.e. without markers) to complexity (with markers). Also, languages with numeral classifiers will not normally change into languages with plural markers and vice versa. Still, a numeral classifier language may change into a plural marking language. In that case, the language will change to a stage of [-PL/-CLF] first before moving further to [+PL/-CLF] as evidenced by Chamorro (Austronesian, Guam), a language which underwent a typological change after contact with Spanish and English (Sanches and Slobin 1973: 12).

This research paper has been regarded as very important in typological studies on numeral classifier systems and number systems, as it is pioneering work proposing the typological relation between the two seemingly unrelated grammatical systems. Although the work concentrates on morpho-syntactic typology, the findings have also sparked the issue of the *semantic* typology of nouns with regard to number in subsequent works such as Greenberg (1972, to be reviewed in Chapter 5)² and Rijkhoff (2002, to be reviewed in §2.2.2). Although currently, due to the advancement of typological studies, more counterexamples against the hypothesis have been observed, the hypothesis remains justified in general, statistically. The languages in which the two systems are

² According to Greenberg (1972: 14), the hypothesis of the complementary distribution of plural markers and numeral classifiers was first proposed in an unpublished paper by Mary Sanches in 1971. The paper was developed and published in the working paper later in 1973.

claimed to co-occur, in fact often show that one of the two systems is optional or used restrictively.

However, since the paper concentrates on the typology of the two grammatical features, the historical description is touched rather briefly. The claim that numeral classifier languages and plural marking languages tend not to change back to [-PL/-CLF] may require more evidence to confirm it, as it appears that the two systems have declined in many languages. The declines of the classifier systems and number systems are described in Aikhenvald (2000:381) and Corbett (2000: 268-271) respectively, for instance. Besides, languages probably do not change in a unidirectional fashion with regard to CNNCs. That is to say, languages with complexity can change into simpler forms as well. Evidence for this possibility may be seen in the diachronic development of CNNCs, as will be illustrated in the later chapters of this thesis.

2.2.2 Rijkhoff's 'Seinsarten'

A typological survey of CNNC_{NSG} appeared in Rijkhoff's (2002) *The Noun Phrase*. The volume aims to discuss various aspects of noun phrases including quantification from typological perspectives. The survey is based on a representative sample of 52 languages across maximal genetic groupings, that is, every phylum (i.e. the genetic grouping at the highest level) is represented by at least one language. According to Rijkhoff (2002: 30-31), only three types of CNNCs are clearly found, namely {N,NUM}, {N,NUM,NSG} and {N,NUM,CLF}. Among these, {N,NUM} is the most common, {N,NUM,NSG} is less common and {N,NUM,CLF} is the least common. Also, languages may employ more than one type of CNNCs. For example, in Tamil, {N,NUM,NSG} is used for human nouns and {N,NUM} for non-human nouns. This suggests that a variety of CNNCs in a single language should be expected. Yet, the dominance of the alternating types is not considered in this survey.

Rijkhoff's main argument regarding CNNCs, however, is that the structural types of CNNCs are associated with the types of nouns (referred to as *nominal subcategories* or *Seinsarten*). Based on morpho-syntactic and semantic properties, nouns are categorized into 6 types, namely *singular object nouns*, *set nouns*, *sort nouns*, *general*

nouns, *collective nouns*, and *mass nouns* (for detailed description on the nominal subcategories, see Rijkhoff 2003: 28-59). These types of nouns are categorized with respect to two semantic features, namely *Shape* and *Homogeneity*.

Shape has to do with whether the referents of nouns are viewed as being spatially bounded. *Homogeneity* deals with whether the referents of nouns are viewed as undergoing a change in form (not size or weight) when added or reduced. To illustrate, in English, the noun *table* has a definite boundary. Also, when a part of it (e.g. a leg) is added to or taken away from it, the original form of *table* changes—in other words, it is indivisible. The noun *table* is therefore characterized as (+Shape, -Homogeneity), corresponding to *singular object noun*. On the contrary, the noun *water* does not have a definite boundary, since its shape can change according to its container. Also when some water is added to or removed from it, the original form of water remains the same. So, the noun *water* is characterized as (-Shape, +Homogeneity), corresponding to *mass noun*.

The nominal categorization is established on the assumption that *nouns* that represent actual physical entities (referents) in the real world are a product of *mental* activity. They are created through a speaker's construal (interpretation) of real objects. So, they have no direct relationship with the physical entities and hence, the nouns and the physical entities are not *perfectly* mapped. As Rijkhoff (2002: 44) puts it, "referents of NPs [Noun Phrases] are not objects in the real world, rather mental constructs that are created, stored, and retrieved in the minds of speech participants".

For that reason, the referents in a single language may be construed differently. For example, the referents of *table* and *water* in English would be construed differently by English native speakers in the manner illustrated above. This fact is reflected in the different quantificational structural patterns when the two nouns are quantified. The noun *table* can be combined with the quantifier directly (for example, *one table*), whereas the noun *water* requires a *unit counter* such as *glass* when it is quantified (for example, *two/many glasses of water*).

For the same reason, speakers from different languages may construe the same real object in a different way (Rijkhoff 2002: 55). The nouns representing the referent

table in English and in Thai may be different in terms of the semantic features *Shape* and *Homogeneity* despite the fact that they refer to the same physical entity. With a Seinsarten-account, one would argue that in English, the referent *table* is construed as spatially bounded (+Shape) and non-homogeneous (-Homogeneity), corresponding to what Rijkhoff refers to as a *singular object noun*. On the contrary, in Thai, the referent *table* is construed as *not* having a boundary in the spatial dimension (-Shape) and as homogeneous (+Homogeneity). Thus, the noun *table* which is categorized as a *singular object noun* in English will be categorized as a *sort noun* in Thai, as reflected in the fact that different types of CNNCs are chosen in the two languages. In English, the noun *table* can be in a direct construction with quantifiers, and plural markers are required if the number of the noun is greater than *one* as in *two/many tables*. In other words, the noun *table* in English requires the structural patterns of {N,NUM} and {N,NUM,NSG} for CNNC_{SG} and CNNC_{NSG} respectively. On the contrary, in Thai, the noun corresponding to *table* requires a numeral classifier both in CNNC_{SG} and CNNC_{NSG} when numerated, as already shown in (2.5). However, the different use of nominal subcategories between the speakers of the two languages involves linguistic knowledge (i.e. knowledge about lexical semantic of the noun); not ontological knowledge (i.e. knowledge about the natural world), as Rijkhoff (2002: 55) notes:

This does not necessarily mean that the speakers of e.g. Thai [...] do not *know* that a table in the physical world is a discrete object, but rather that this particular piece of knowledge is simply not part of the lexical semantics of the noun (instead it would be part of the encyclopaedic knowledge about the referent of NP (*italics original*))

That is, the speakers of numeral classifier languages know that the referents of *table* and *water* are physically different (i.e. water is not discrete, whereas a table is), but the physical characteristics as such are ignored in the lexical meaning of the nouns.

To sum up, according to Rijkhoff's survey, three major types of CNNC_{NSG} are observed along with their frequencies. We have also learned that a variety of CNNCs can be expected in a single language. Nevertheless, diachronic issues do not receive discussion in Rijkhoff's volume. The main argument is focused on the idea of

Seinsarten, suggesting that the differences in structural patterns of CNNCs in a single language and in different languages can be explained by semantic properties of nouns. So, an implication from the study is that the diversity of CNNCs is associated with the differences in the construal of referents of nouns in languages. In other words, semantic typology governs morpho-syntactic typology. The idea of *Seinsarten* is useful and interesting because it provides universal criteria for categorizing nouns with regard to quantification across languages. Also, it sheds some light on the reason why there is such a variety in CNNCs. However, since the six nominal subcategories are established based on the structural patterns (or morpho-syntactic properties) in CNNCs, circularity would be impossible to avoid if we claimed that CNNCs are different due to the difference in the nouns used in the constructions.

2.2.3 Hurford's 'The interaction between numerals and nouns'

Hurford's (2003) '*The interaction between numerals and nouns*' is a typological survey of morphosyntactic phenomena occurring within numeral-noun constructions including *cardinal*, *ordinal* and *plural numerals* (to be defined later), focusing on "syntagmatic distribution and the internal structure of numerals" (2003: 561). The survey is based on a sample of 25 languages that were primarily selected on the basis of maximal geographical distance and genetic diversity, though with some concentration on European languages.

Regarding CNNCs, this paper brings up various other structural types of $CNNC_{NSG}$ and related constructions than those previously reported. Illustrated below are {N,NUM,OBL,NSG} and {N,NUM,CLF,NSG,INDEF} as observed in Welsh (2.8) and Sinhala (Indo-European; Sri Lanka) (2.9) respectively. The two types are claimed to be rare.

(2.8) *Welsh (Hurford 2003: 569)*
naw o ddynion
nine of man.PL
'nine men'

(2.9) *Sinhala* (Hurford 2003: 574)

minissu *tun* *den-ek*
man.PL *three* *CLF* (lit. 'people')-INDEF
'three men'

In addition, in relation to {N,NUM,CLF}, the paper argues that although the type is not typical in European languages, there exist numeral classifier-like constructions found in certain European languages, for example, English *forty head of cattle*, where the noun *head* functions like a numeral classifier. Yet, unlike a typical numeral classifier construction, the preposition *of* is required in the construction. Also, in some languages, such as Bulgarian, Hungarian and Scottish Gaelic, there is a special set of numerals used with nouns denoting humans. The use of special sets of numerals counts as examples of numeral classifier-like constructions (Hurford 2003: 570-571).

(2.10) *Scottish Gaelic* (Hurford 2003: 571)

ochdnar *chloinne*
eight.CLF *children*
'eight children'

There are three other numeral-noun constructions which deserve a mention. These numeral-noun constructions are structurally identical to CNNCs, but owing to the difference in constructional meanings, they are not actually regarded as CNNCs. These constructions reflect that the situation for morpho-syntactic typological studies at least of numeral-noun constructions, the constructional meaning cannot be ignored. The first such construction is the numeral-noun construction consisting of a plural numeral and a plural noun as in Finnish in (2.11).

(2.11) *Finnish* (Jensen 1934: 47 in Hurford 2003: 586)

kahdet *sukat*
two.PL *sock.PL*
'two pairs of socks'

The construction is structurally identical to {N,NUM,NSG,NSG} where {NSG} is an agreement. However, in this case, the meaning of the construction containing plural numerals is different from CNNCs. In plural numeral constructions, the constructional meaning is ‘*n groups of*’ (2003: 585) (like *two groups/pairs of socks*). According to Hurford, the plural numerals are found in Finnish, Estonian and Icelandic but nowhere else in his sample.

Another construction which is mentioned but is not further considered in the survey is the *ordinal numeral-noun construction* where in the same language the ordinal is not different from the cardinal numeral, for example English *one day* (cardinal numeral-noun) versus *day one* (ordinal-noun) (for another example, see footnote 1 above). Thus structurally, it is identical to {N,NUM} of CNNCs. The difference between the two constructions lies in numeral-noun word order. In this case, the one which expresses the order of things does not count as a CNNC.

The final construction to be mentioned is the *definite* cardinal numeral-noun construction (comparable to English *the two books*). This construction is found in Sinhala where the unmarked cardinal numeral-noun construction expresses the meaning *definiteness* of the whole noun phrase. For example,

(2.12) *Sinhala* (Hurford 2003: 577)

pot deka
book two
 ‘*the two books*’

(2.13) *Sinhala* (Hurford 2003: 577)

pot deka-k
book two-INDEF
 ‘*two books*’

The examples as shown above may cause a difficult problem: the ambiguity as to whether or not the indefinite marker should count as an extra element specifically required in CNNCs in this language (comparable to number markers or numeral

classifiers, for instance). In this thesis, some morphemes such as gender markers in French are not considered as (required) extra elements for CNNCs because they have no function relevant to number at all; the markers as such are seen in CNNCs just because of the structure rule of the noun phrase. Although the indefinite marker in Sinhala is not associated with number, the constructional meaning will change to definiteness without the definite marker, as shown in (2.12). In this case, the indefinite marker should be considered as a required extra element in Sinhala. Therefore, numeral-noun constructions without the *indefinite marker* in Sinhala, namely {N,NUM} and {N,NUM,CLF,NSG} are not CNNC_{NSG}, as they include the meaning of definiteness in their constructions.

2.2.4 Conclusion

CNNCs, especially CNNC_{SG}, are constructions that have attracted little attention in the general linguistic literature. According to the previous studies on CNNCs just described above, certain structural types of CNNC_{NSG} have been documented; notable recent surveys include Rijkhoff (2002) and Hurford (2003). Still, these types of CNNCs require further analysis of their distribution and frequencies in the world's languages. Besides, it is implicitly stated in previous research that a variety of types of CNNCs can be observed even in a single language, but the issue of predominance of types has not yet been considered. Although there has been typological research into CNNC_{NSG}, apart from the work by Sanches and Slobin (1972), little has been done on the diachronic development of CNNC_{NSG}. The diachronic studies can perhaps give a valuable account of the variety of structural patterns of CNNCs in modern languages. After all, the earlier research has provided the background knowledge for understanding the current situation of CNNCs and this is necessary for designing the current project properly.

3 How to Trace the Evolution of CNNCs

There are three steps in tracing the evolution of CNNCs. First of all, structural types of CNNCs across languages are explored to see the possible types attested in living languages (§3.1). Then, historical origins of structural types of CNNCs are examined, using old written records along with theoretical considerations (§3.2). Finally, the historical paths of structural types of CNNCs taken from the second step are combined into globally evolutionary structures (§3.3).

3.1 Exploring types of CNNCs

3.1.1 *Sampling*

The first step in investigating the evolution of CNNCs is to find the range of possibilities of structural types of CNNCs in modern human languages. Ideally, to be completely knowledgeable about all the possible structural types of CNNCs, one would have to examine all the roughly 6,000 living languages. This is obviously unfeasible because the majority of them are not well documented at all. Although there are hundreds of languages already well described, it would still be logistically quite difficult to examine all of these languages. Therefore, we need a method commonly known as *language sampling*—a method of using a representative sample of the documented languages in studying cross-linguistic variation (Croft 2003:19). In sampling, it is generally assumed that the more diverse genealogically and geographically the languages selected are, the more likely it is that one will be able to find the most diverse types of linguistic feature under study. With this assumption, to display the greatest possible variety, at least one representative language from each *language family* (i.e. top-level genetic group of languages) was chosen.

According to preliminary research, however, in some language families (e.g. Indo-European), the structural types of CNNCs seem to be more diverse than others, especially in families which are larger or have greater time depth. The diversity is also

perhaps attributable to the fact that those families are better described than others. In any case, this suggests that in some language families, a single representative language would probably be unable to reflect a realistic picture of the CNNCs in that family. Therefore, rather than selecting only one language representing one family, we selected at least one language representing one *genus* (i.e. a genealogical group of languages with a time depth no greater than 4,000 years, comparable to the Indo-European subfamilies (Dryer 1992: 83-84)), if the genus provided useful information about CNNCs.

The sample also includes representative languages from *creoles and pidgins* (i.e. simplified mixed languages, developed after intensive contact, mostly with European languages), chosen from different areas across the globe, namely North America, South America, Africa, Australia-New Guinea and Southeast Asia, plus some *language isolates* (i.e. languages with no known genetic relationship to other living languages) in which the appropriate sources for CNNCs are available. This procedure yields a total sample size of some 250 languages, a size in which we may find the majority of the possible structural types of CNNCs attested in human languages. A listing of the sample languages representing language groups is given in Appendix 1.

In relation to the selection of the representative languages, it is assumed here that any living languages which are well-documented and provide sufficient information on CNNCs have an equal chance of representing the genus of the family. Besides, the languages were selected without their structural patterns of CNNCs being known in advance. The selection of languages within the genera is thus random. So, an attempt has been made to ensure that the sampling has been conducted by avoiding genealogical and geographical biases as much as possible.

Regarding the language classification, the thesis follows the classification presented in the *World Atlas of Language Structure* (Haspelmath, Dryer, Gil and Comrie 2005; henceforth *WALS*), a large cross-linguistic database for typological studies. In fact, the language list in *WALS* is not a linguistic classification as such; rather it is a sample in its own right. Therefore, it includes only the languages documented in this sample. Even so, it comprises 2,560 languages, almost half of the

world's languages today. There are a couple of reasons why WALS suits the current project. The first is that it is the database which we believe the most up-to-date at the time of the current data collection (September-December 2005). Also in some cases, WALS provides remarks indicating the differences between WALS and previous frequently cited database, namely Ruhlen (1987) and Grimes (2000) in terms of language names and classifications. So, these differences as highlighted in WALS help us to ascertain whether the languages we selected are problematic in any way. Besides, the classification as a *genus* has some advantages over the classification as a *branch* as used in some other sources. That is to say, the notion *genus* has a clearer definition than the notion *branch* does in terms of *time depth*, whereas the term *branch* can mean either a subfamily or a genus. To compare like with like in terms of time depth, WALS was chosen because it is stated clearly whether the languages selected are in the same genus. This proved necessary when considering statistical frequencies, since counting genera is less error-prone than counting languages or families. Furthermore, although some languages in the WALS sample do not provide useful information on CNNCs at all, a great number of languages in the sample provide the information with regard to *word order of noun and numeral* (Dryer 2005b), which is useful for CNNCs. The information provided always includes bibliographic references with page citations. This proved opportune and helped speed up the process of data collection, as we could ensure that the languages chosen were likely to provide the relevant information about CNNCs.

3.1.2 Data sources

The language data reported in the thesis are mainly based on the *survey of reference grammars*, a methodology for extracting grammatical information from published reference material (Bybee, Pagliuca and Perkins 1994: 35). Since we need to use a large database for the phenomenon in question, this methodology is thus more practical than eliciting information from native speakers through questionnaires or from texts. Also, the phenomenon is relatively well-described in most grammars, so the data can be obtained directly from grammars. The methodology, however, was supplemented by samples from texts (where glosses were provided) as well as from previous literature.

Also, where necessary, the information was checked with native speakers or language experts. However, as noted in Bybee, Pagliuca and Perkins (1994: 35), a weakness of the methodology involves the comparability of information. That is, different authors may use different frameworks or terms for the same linguistic phenomenon. For example, some authors may refer to every instance of number distinction in nouns like *non-numeral quantifiers* (e.g. *many*) as *number markers*. This mistakenly suggests that the language has a number marking system. The problem is more obvious in reference material written several decades ago when the standard grammatical analysis was underdeveloped.

The grammar reference survey designed for the thesis is aimed at investigating CNNCs not only with regard to structural patterns of CNNCs, but also related issues necessary or useful for tracing back the evolutionary paths of CNNCs. These issues include (1) frequencies, distribution and mode of predominance of types of CNNCs, (2) evolution of number-marking systems, (3) the polygenesis of numeral classifier systems, (4) the correlation between types of CNNCs and noun class systems (henceforth *the correlation*), and (5) quantificational expressions before the rise of CNNCs. The survey format (cf. Appendix 3) consists of two parts as follows.

Part I deals with the necessary background information of the language, namely language name, its affiliation (at two levels, i.e. family and genus), the region where the language is spoken, and all references for the language data. The affiliation and region are required for statistical purposes and for certain matters pertaining to the geographical maps. As for the background information, the thesis follows the information given in *WALS*.

Part II is designed for extracting grammatical information from descriptive grammars. This part is divided into 3 sections dealing with structural types of CNNCs, number systems, and classifier systems. The questions in Section 2 and Section 3 are taken from *Lingua Descriptive Studies: Questionnaire* (Comrie and Smith 1977).

Section 1 consists of 4 questions (Question 1 – Question 4). The objectives of this section are to reveal structural types of CNNCs across languages and to see whether

they are used as the *basic type* (i.e. types used generally; alias *primary* or *dominant*) or *subsidiary type* (i.e. types used restrictively) in a given language.

Question 1: Does the language have CNNC_{SG}? If so, illustrate the structural patterns of CNNC_{SG}.

Question 2: Are the structural types observed in Question 1 used with or without constraints (e.g. used with a small set of nouns or a particular context)?

Question 3: Does the language have CNNC_{NSG}? If so, illustrate the structural patterns of CNNC_{NSG}.

Question 4: Are the structural types observed in Question 3 used with or without constraints (e.g. used with a small set of nouns or a particular context)?

For Questions 1 and 3, all structural patterns of CNNCs observed in the referential material were collected (for the operational definitions of CNNCs and the extra elements used in CNNCs, see §2.1 and Chapter 5 respectively). The information obtained from Questions 1 and 3 is analyzed and presented in Chapter 6. As for Questions 2 and 4, these questions are aimed at revealing the basic types and subsidiary types of CNNCs. This information is required for classifying *language types* (i.e. classifications on the basis of basic types), which in turn is necessary for statistical purposes and for making maps illustrating geographical distribution of types of CNNCs. The answers obtained from Questions 2 and 4 are analyzed and presented in Chapter 6 and Chapter 7.

Section 2 consists of 3 questions (Question 5 - Question 7). The objective of this section is to disclose information about the number-marking system in languages. This information is useful for the issue of the evolution of number-marking systems (cf. §5.1.3).

Question 5: “Does the language have a number-marking system in nouns? [and if so] Do different classes of nouns behave differently in this respect (e.g. animate versus inanimate)?” (Comrie and Smith 1977: 34).

Question 6: “Is the system of marking number obligatory or optional?” (Comrie and Smith 1977: 34).

Question 7: “If the language has no system of number-marking in the noun, does it have other means of indicating number, such as the use of a word meaning ‘many’, etc.?” (Comrie and Smith 1977: 34).

For Question 5, the reader is referred to §5.1.1 for the operational definition of the number-marking system in nouns. The number-marking system may have been explicitly shown in the answer to Question 1 (e.g. English *two dogs*). However, in some languages, such as Hungarian, the existing number markers are not employed in CNNCs. If we looked only at the information in Question 1, we might misunderstand that Hungarian does not have a number-marking system. Also, some classes of nouns may be marked for number, whereas others may not be. Question 6 is required because the number-marking system in nouns in some languages is obligatory, but optional in others. For the Questions 5 and 6 the data are taken mainly from WALS. The answers obtained from these two questions are analyzed and presented in Chapter 7. Question 7 is required because it is assumed that non-numeral quantificational expressions are conjectured to be a quantificational stage before the rise of CNNCs. Therefore, we need to know the means of indicating number in the languages in which number-marking systems are not available. The answers obtained from Question 7 are analyzed and presented in Chapter 4.

Section 3 consists of 2 questions (Question 8 - Question 9) concerning classifier systems, namely *noun classes* (or *genders*) and *numeral classifiers* (for the operational definitions of the three terms, see §5.2). The objective of this section is to reveal the classifier systems across languages. The information obtained from Question 8 is required for the issue of correlation, while the information obtained from Question 9 is required for the issue of polygenesis of numeral classifier systems. The answers obtained from Question 8 and Question 9 are analyzed and presented in Chapter 7 and Chapter 6 respectively. The two questions are straightforward, as follows:

Question 8: “Are nouns divided into classes or *genders*?” (Comrie and Smith 1977: 34).

Question 9: “Does the language have numeral classifiers?” (Comrie and Smith 1977: 35). If so, is the system of numeral classifiers obligatory to all nouns?

For Questions 8 and 9, the answers were partly taken from WALS. The answer to Question 9 may be obtained from Question 1, however, in some languages, such as Persian (Indo-European; Iran) and Turkish (Altaic; Turkey), the numeral classifiers are employed (and perhaps optionally) only with some nouns. The reader is referred to Appendix 3 to see the format of this reference grammar survey designed based on the nine questions above. Also, in the introductory pages of Appendix 3, the thesis provides a description on what should be inferred from various types of answers to the questions.

3.2 Investigating historical origins of CNNCs

After exploring possible structural types of CNNCs in the world's languages, the next step is to trace the historical development of these structural types of CNNCs so as to understand how these patterns came into being in the world's languages. The investigation is performed by using *evidence from old texts* as well as *theoretical diachronic approaches*.

3.2.1 Evidence from old texts

Old written records are very helpful direct evidence for tracing the possible historical routes of CNNCs. This is because old documents are the place where the previous structures of CNNCs can be observed. Then we can compare them with more current patterns to observe the change, and sometimes even the factors which contribute to the change as well. In this study, the old texts are obtained from previous historical studies relevant to the change of CNNCs in a particular language or a language group. The old texts themselves come from Old Chinese, Old and Middle English, Classical Arabic, Classical Tamil, Old and Middle Khmer, among other languages.

Although the old texts provide concrete evidence, those which present information directly involving the change in CNNCs are very rare. In addition, some written records are not old enough to reflect a change in CNNCs.

In addition to the problem of rarity and shallow time depth, there is another problem arising from using this kind of evidence. Namely, the old written records may not perfectly reflect the spoken form of the language in those days. Alternating patterns may have existed even if they are not recorded. Present-day written Khmer provides a good example. CNNCs in modern written Khmer are normally {N,NUM,CLF} but in colloquial speech, CNNCs can be both {N,NUM,CLF} and {N,NUM}, as shown in the examples (3.1 a-b) (Jacob 1965: 145). Supposing there was no other type of evidence, one might mistakenly conclude that the pattern of {N,NUM,CLF}, as the only pattern recorded, was the primary pattern of CNNCs in this period.

(3.1) *Modern Khmer (Jacob 1965: 145)*

<i>a</i>	<i>ko:n</i>	<i>bɔy</i>	
	<i>child</i>	<i>three</i>	
	<i>'three children' (Colloquial)</i>		
<i>b</i>	<i>ko:n</i>	<i>pi:(r)</i>	<i>néək</i>
	<i>child</i>	<i>two</i>	<i>CLF (lit. 'person')</i>
	<i>'two children' (Written)</i>		

3.2.2 Theoretical diachronic approaches

Due to the limitations of the evidence from old texts, we need to use other types of evidence in investigating the historical developments of CNNCs. These types of evidence (which in fact are theoretical approaches to diachronic studies) will be referred to as *theoretical approaches* here. The relevant theoretical approaches include (a) Grammaticalization, (b) Lexicalization, (c) Internal Reconstruction (IR) and the Comparative Method (CM), and (d) Language Contact. These theoretical approaches will be reviewed and discussed in detail in §3.2.2.1-3.2.2.4 below. Although the theoretical approaches employed in the investigation of the history of CNNCs will inevitably give a conjectural conclusion which is less reliable compared to the historical evidence, such theoretical approaches can be applied to broader cases. These theoretical approaches will become more reliable if they can be used to support each other or together with the historical evidence.

3.2.2.1 Grammaticalization theory

The theory of grammaticalization (alias *grammaticization*) is a theoretical framework for language change offering an explanatory account of how grammatical categories arise and develop. Due to the fact that the historical development of types of CNNCs and the origins of the *quantificational extra elements* (cf. Chapter 5) (which are grammatical categories) lie in grammaticalization processes, in this section the theory of grammaticalization is briefly reviewed.

It is widely assumed that the basic linguistic elements in languages can be divided into two categories, namely *lexical elements* (alias *content words*) and

grammatical elements (alias *function words*). Lexical elements describe things, actions, and qualities. In accordance with this definition, nouns, verbs and adjectives are instances of this category. The lexical elements are in an open-class paradigm, as the size of the class can be increased. On the other hand, grammatical forms serve morpho-syntactic purposes, such as indicating relationships between nouns, linking parts of discourse, referring to referents already identified, and so on. These grammatical functions are represented by prepositions (e.g. *at*), connectives (e.g. *as*, *for*) and pronouns (e.g. *he*) respectively. Grammatical elements also include derivational forms (e.g. *-ity*), clitics (e.g. the genitive form *'s*) and inflections (e.g. *-s* indicating third person singular subject of the verb in the present tense). Generally, the members of the class hardly change, so they are in closed-class paradigms (Hopper and Traugott 2003: 4). However, the division is in fact on a continuum, as the boundary between the two categories is not absolutely clear. There are a number of linguistic items which are difficult to categorize. Among these are numerals in which the lexical meaning remains but a grammatical purpose as a *quantifier* is also served. For this reason, different languages treat numerals differently. Those numerals may belong to the class of nouns in some languages, but may belong to the classes of adjectives, verbs or quantifiers in others (Corver, Doetjes and Zwarts 2007: 752). However, the numerals are normally in a closed-class paradigm, and for this reason, the class of numerals is functional rather than lexical.

Turning to *grammaticalization*, the term was first coined by the French linguist Antoine Meillet who defined it as

[L]e passage d'un mot autonome au rôle d' element grammatical ("the shift of an independent word to the status of a grammatical element") (1912, quoted and translated in McMahon 1994: 160).

Note that *un mot autonome* (an independent word) in fact corresponds to a lexical element, as a lexical element is syntactically independent, whereas a grammatical element must occur in relation to a lexical element. This classic definition can be seen in

general grammaticalization phenomena, such as the development of auxiliaries from verbs. Cross-linguistically for example, the lexical verbs meaning 'want' have developed a grammatical function as a future marker, as in the case of the English auxiliary *will* from Old English verb *wyllan* 'want'.

The term grammaticalization was later given a broader definition by the Polish linguist Jerzy Kuryłowicz (1965: 52):

Grammaticalization consists in the increase of the range of a morpheme advancing from a lexical to a grammatical or from a less grammatical to a more grammatical status [...]

This concept of grammaticalization emphasises the extension of the *range* of grammatical status in linguistic elements. The definition thus includes the development of more grammatical elements from less grammatical elements. An example is the development of the plural marker from the third person plural pronoun (Heine and Kuteva 2002a: 237-8) in many languages. However, in some cases, it is difficult to say whether one element is more grammatical than another. All we can say perhaps is that the grammatical form just assumes a new grammatical function.

In the recent literature on grammaticalization, Hopper and Traugott (2003) hold that the grammaticalization process not only takes place at a lexical level but also in constructions, such as English *ƒa hwile ƒe* 'that time that' being grammaticalized to *while* (a temporal connective). That is, the meaning of the larger construction is fused in a noun, *hwile* 'time' (Hopper and Traugott 2003: 4).

Furthermore, it is generally realized that grammatical functions may develop from pragmatic meaning and prosodic features instead of lexical meaning. For example, word orders which pragmatically signal *theme* (i.e. the first part of an utterance, functioning as the beginning of the message) and *rheme* (i.e. the second part of an utterance, expressing the information about the *theme*) can change into grammatical word-orders signaling syntactic cases for subject and object respectively. Note that in this case there is no lexical meaning involved. Furthermore, phonological or prosodic

features may also assume grammatical functions, such as a particular tone functioning as a numeral classifier in Beijing Mandarin (Tao 2005, Wischer 2006: 129) (cf. §8.1.2).

To sum up, it can be observed that the concept of grammaticalization has become broader and broader, so as to cover new phenomena of grammatical changes observed. In any case, the central concept of grammaticalization lies in the origins and development of grammatical elements. The important question is how grammaticalization processes happen.

It has recently been emphasized that the mechanisms of grammaticalization may be invoked when a particular lexeme which has potential source meaning is in the proper context or construction (e.g. Bybee 2003: 602-603). A good example of grammaticalization processes can be found in the grammaticalization of *be going to/be gonna* (like *She's going to die.*) in English. The illustration and examples here are mainly based on the analysis in Hopper and Traugott (2003: 2-3) and a few terms in Wischer (2006: 131-132) as follows.

(1) Originally, *be going to* is used as a lexical motion verb in progressive form with a locative adverb such as *I'm going to London*. At this stage grammaticalization does not yet occur. However, note that the meaning 'go to' (which can be used both to express a goal and an aim) has the potential to grammaticalize.

(2) *Be going to* is used with the purposive directional construction, as in *I'm going to London to marry Bill*. In this context, *be going to* is still used as a lexical verb, but notice that the purposive context may trigger the change in meaning easily.

(3) The locative adverb can be omitted, that is, *I'm going to marry Bill*. At this stage, the first mechanism, namely *pragmatic inferencing* (i.e. deducing the new meaning of the linguistic items going to be grammaticalized in a given context (Wischer 2006: 129)) comes into operation. This is because the purposive construction implies futurity. When the directional phrase (i.e. *to London*) is omitted, the futurity will become salient. There are two types of pragmatic inferencing. The first is *metaphor* and the second is *metonymy*. The former involves the process whereby the concept of "a more basic or concrete entity" is transferred to a more abstract one" (Wischer 2006: 131).

The latter has to do with the contiguity of words in juxtaposition (Wischer 2006: 131). In this case, *be going to* can assume the meaning of futurity as a result of metonymy. That is to say, the verb phrase *be going to* followed by the location conveys the meaning of 'movement in progress'. Once there is a verb or an action following the phrase *be going to*, then the implicature 'movement in progress' (implying *the futurity*) is carried on to an action (e.g. *she is going to open the window.*). At this stage, frequency of use may play an important role, this implicature of futurity became a meaning of *be going to* (cf. Bybee 2003; Wischer 2006: 131).

(4) The simultaneous process referred to as *reanalysis* happens, that is, *be going to* is reanalyzed. The preposition *to* 'purposive marker' is combined tightly with *be going to*, leading to the new grammatical construct used as an auxiliary expressing immediate future. The reanalysis can be written as: [I am going [to marry Bill]] (with 'purposive' meaning) is reanalyzed as [I [am going to] marry Bill] (with 'immediate future' meaning). At this stage, we can see that the lexical meaning of 'go' has been bleached (Hopper and Traugott 2003: 3).

(5) The new construct *be going to* is extended in its use to more general contexts, that is, including those which may not be compatible with purposive meaning (for example, *I'm going to like Bill*). This is generally known as *context generalization* (Heine and Kuteva 2002a: 2)

(6) Like a typical grammatical form, *be going to* undergoes further changes, such as phonetic reduction (or *erosion*), i.e. *be gonna*.

(7) The stages of the grammaticalization of *be going to* co-exist in modern English. The co-existence such as this is referred to as *divergence* and also as *layering* (Hopper and Traugott 2003: 118; Wischer 2006: 132).

3.2.2.2 Lexicalization

Lexicalization in diachronic linguistics generally refers to the phenomenon involving the origins and development of new lexical items in a language. Brinton and Traugott's (2005) *Lexicalization and Language Change* offers a comprehensive review of the literature on lexicalization and a detailed discussion on this subject. In the current

project, there are certain language changes associated with lexicalization, namely the origins of numerals, the fusion of nouns and number markers or noun class affixes, and the fusion of numerals and numeral classifiers. In this section, a general introduction on lexicalization will be given and the issues pertinent to the phenomena observed in the project will be briefly reviewed along with examples taken from this source.

Like grammaticalization, the term has been used in various senses. In the very broad sense, some types of *word-formation* are typically treated as lexicalization. An example is *compounding*, which refers to the unifying of independent words, such as the formation of *blackboard* from *black* and *board*. Another instance is *clipping*, which refers to the omission of syllable(s) in multisyllabic words, such as the formation of *flu* from *influenza*. These types of word-formation are crucially treated as lexicalization because the morphemic boundaries of the unified independent lexemes have been lost, giving rise to a single morpheme. The word-formation which involves a shift from grammatical to lexical functions is also widely recognized as lexicalization. For example, the verb (*to*) *off* (in *to off the microwave*) is derived from the preposition or adverb *off* (in *turn off*). The shift in syntactic categories such as this is known as *conversion*.

Lexicalization is also characterized as the process of fusion of linguistic items, giving rise to a decrease in compositionality and hence a lexicalized item. Some of the phenomena observed in this thesis involve lexicalization in this sense. A few types of fusion can be described as follows. The first and one of the commonest concepts of lexicalization is the phenomenon in which lexemes are derived from syntactic phrases or constructions. The phenomenon is referred to as *univerbation*. For example, the lexeme *handicap* is developed from the noun phrase *hand in the cap*, and the lexeme *more or less* 'somewhat' used to be a phrase. The former example shows a phonological reduction, suggesting that it is older and hence rather semantically opaque, whereas the latter is more recent and hence relatively transparent. In more marginal cases, a whole utterance can be compressed into a single lexeme, such as in *goodbye*, which is derived from *God be with you*.

The process of fusion also involves the phenomenon where a simple lexeme has developed from a complex lexeme. The complex lexeme was once formed by a process of word formation, but now is difficult to analyse and hence not recognizable due to semantic, morphological, or phonological reasons (Bauer 1978: 6 in Brinton and Traugott 2005: 50). The current lexicalized form is synchronically unanalyzable and is probably semantically opaque. For example, the current lexeme *mermaid* developed from the Old English complex lexeme *mere* 'sea' + *mægd(en)* 'maiden'. Once the etymological meaning of each constituent of the word becomes unrecognisable among general users, the word can be regarded as lexicalized. Besides, the derivational morpheme which was once a constituent of the compounded word can also be considered as an instance of lexicalization if the derivational morpheme becomes unproductive—that is, the users are hardly aware of them and hence the fused morphemes are likely to be perceived as a single lexeme. Examples are the suffixes *-ric* and *-dom* which are derived from Old English *rice* 'realm' and *dom* 'setting, jurisdiction' respectively. These two suffixes are relatively unproductive compared to productive ones such as *-ly*, and *-ment*.

In addition, another characterization of lexicalization as fusion is concerned with the shift from morphological elements into phonological or syntactic elements. This phenomenon is called *demorphologization* or *demorphemization*. For example, the current lexeme *alone* is derived from the construction consisting of *all* and *one*. In this case, the form of the morpheme *all* has been almost lost, and only the original phonological element (i.e. *al*) remains as an indistinguishable part of the word, while the grammatical function and meaning of the word is fused in the new word *alone*. Demorphologization also includes the fusion of bound morphemes into the root. For example, the word *whilst* is derived from the construction consisting of *while-s-t*, where genitival *-s* and excrement *-t* are bound morphemes which are integrated into the root *while*. In this case, the morphemes have lost their grammatical-semantic role to the new word, and only the phonological element remains as an unanalyzable and opaque part of the word. Note that although *while* belongs to the grammatical category, the change itself is treated as lexicalization. As for the case of *syntacization*, this can happen when

a morphological element changes its status into more independent elements, such as a clitic or a word. For example, the autonomous words like *ology* ‘subject of study’ and *ism* ‘doctrine, theory, practice’ developed from the derivational suffixes *-logy* (as in *zoology*, *sociology*) and *-ism* (as in *socialism*, *communism*). These forms can also be regarded as *clipping*.

The description above is a partial account of lexicalization. For a comprehensive discussion of this subject matter, the reader is referred to Brinton and Traugott (2005).

3.2.2.3 Linguistic reconstruction

Linguistic reconstruction is an attempt to recover linguistic prehistory or remote past characteristics of a single language or a language family based on actual synchronic data, whether earliest written records or spoken language. The methods which are used to reconstruct a *pre-language* (i.e. a hypothesized past language before the appearance of the current language) and a *proto-language* (i.e. a hypothesized past language before the appearance of the current language family) are generally known as *Internal Reconstruction* (IR) and the *Comparative Method* (CM) (alias *Comparative Reconstruction*) respectively (McMahon 1994: 6; Fox 1995: 1-6).

IR and CM can be applied to several areas of grammar, namely the lexicon, phonology, morphology and syntax. Since the current project deals with CNNCs which are syntactic phenomena, however, this review will concentrate on this area of grammar. Although it is widely recognized that syntactic reconstruction seems to be problematic, this does not mean that it is hopeless. Such methods may be helpful for the current study in some cases. The two methods can therefore be employed to suggest hypotheses. We may then see whether the hypotheses are consistent with the findings obtained from other approaches.

On the basis of IR, the historical linguist exploits *alternations* (i.e. alternating linguistic forms or patterns descended from the same source) within a single language as evidence to hypothesize the previous stages of current forms. Therefore, the method of IR is based on the assumption that the alternating forms have descended from a single

source in the pre-language³, and these alternating forms reflect “the continuity of language change” (Fox 1995: 147). In other words, the method of IR will be possible only when alternations in the current language can be identified (Fox 1995: 146-147, 190). Note that the alternations between forms are systematic, resulting from regular changes (Fox 1995: 147), not just one or two instances occurring accidentally. The alternations which occur systematically are known as *correspondences*. IR is successful in phonological and morphological reconstruction but not in syntactic reconstruction. This is because the phonological or morphological alternations are historically related and can be explained in terms of regular change such as sound change. So, we can find a “continuity between successive forms” which represents the current alternating forms. On the contrary, the alternating syntactic structures are not historically related in that way. Syntactic change “is essentially an analogical process rather than an evolutionary one” (Fox 1995: 190). To illustrate, consider examples of phonological alternations and the continuity between successive forms (Fox 1995: 154-155).

<i>knife</i>	-	<i>knive(-s)</i>
<i>life</i>	-	<i>live(-s)</i>
<i>wife</i>	-	<i>wive(-s)</i>
<i>loaf</i>	-	<i>loave(-s)</i>
<i>calf</i>	-	<i>calve(-s)</i>

By means of IR, we may set up the correspondence and establish pre-phonemes as follows. Based on the data above, *f* as the final consonant is regularly changed into *v* when followed by *-es*. So, the rule of sound change may be set up and represented by the notation as $f \rightarrow v/_es$. Since the change occurs systematically in a particular phonological context, *f* and *v* are treated as the alternating forms of the same phoneme. Also, a correspondence set for the alternating forms can be established as $f \sim v$. The

³ However, one may question whether it is possible that these alternations have always existed in the language. Historical linguists can avoid this tricky issue by making their statements implicational, like “the alternant sounds X and Y go back to, say, X, *if* they go back to a non-alternating stage at all” (Miranda 1975: 292; *italics mine*). In any case, according to evidence from a long recorded history, it is found that alternations are likely to decline and have an end (Miranda 1975: 292).

sound change $f \rightarrow v/_es$ is also treated as a historical account for the continuity of the successive forms f and v , where f is a reconstructed form.

Turning to syntactic reconstruction, although there are syntactic alternative forms, they may be pragmatically or stylistically different. For example, it is difficult to conclude which forms are earlier (at least in the same way as with phonological reconstruction) between *I go there often* and *I often go there* (Fox 1995: 190).

However, according to Fox (1995: 190-191), there are some particular principles used in IR for syntactic reconstruction if syntactic alternations can be identified. The first is the principle of *simplicity* which assumes that the earlier form might have been simpler than the current forms. Consider the two sets of the interrogative sentences (right column) below:

Set I	<i>He is in London.</i>	<i>Is he in London?</i>
	<i>She has seen the film.</i>	<i>Has she seen the film?</i>
	<i>They will come next week.</i>	<i>Will they come next week?</i>
Set II.	<i>He lives in London</i>	<i>Does he live in London?</i>
	<i>She has brown eyes.</i>	<i>Does she have brown eyes?</i>
	<i>They want to come next week.</i>	<i>Do they want to come next week?</i>

(Fox 1995:190)

Based on the data above, we can set up a correspondence set of interrogative sentences as Verb+Subject+X (X = the remainder of the sentence) (Set I) and Auxiliary+Subject+Verb+X (Set II). Set I is used when the finite verb is the verb *be* or an auxiliary, such as *has* or *will*. On the other hand, Set II is used when the finite verb is not the verb *be* or an auxiliary. In this case, the verb *do* which is an auxiliary verb must be added. By means of the principle of simplicity, the former alternant is supposed to be original, since it is simpler (Fox 1995: 191).

However, it is also noted that one problem with this principle is that in cases where the number of the constituents is the same, such as X+Verb+Y versus X+Y+Verb, the principle of simplicity may not be helpful. The solution is that we have to identify a *marked* structure and an *unmarked* structure. The marked structure is a structure which

is restricted in use or is used with special meaning, whereas the unmarked or basic structure is used in more general contexts and more frequently observed. By means of IR, the unmarked form is considered to be an earlier form from which the marked form developed. For example, in Latin, the verb initial position is marked, whereas the verb final position is unmarked. Therefore, it can be conjectured that the verb initial position developed from the verb final position (Fox 1995: 199-200).

Another principle of reconstruction is based on typological generalizations or the *typological approach to syntactic reconstruction* (Thomason 2006: 398). This approach assumes that the typological generalizations are supposed to be valid across time. For example, languages possessing NAdj word order (i.e. nouns precede adjectives) tend to have VO word order (i.e. verbs precede objects) (Thomason 2006: 398). By means of IR, if Language X is a language with NAdj word order and has two alternating structures, namely VO and OV (i.e. objects precede verbs), it would be predicted that the original type is VO. The principle of the typologically-based approach is that the pattern which is consistent with the typological generalizations is expected to be original, while the other which is inconsistent with the generalizations is likely to be more current.

Following the typological approach to syntactic reconstruction, if we can find a typological correlation between types of CNNCs and other linguistic features, we may be able to specify the earlier type and the more current one. But as noted in Fox (1995: 252), the typological generalizations are mostly not absolute but must rather be regarded as preferences. In other words, they show only a general tendency, and therefore there are perhaps some exceptions. If the case under study is exceptional, the approach will be useless. Also, in reality, the types of languages are subject to change leading to the *typological inconsistencies* which remained in some languages for many centuries (Watkins 1976 in Thomason 2006: 398).

So far it seems likely that syntactic reconstruction by means of IR is limited in use. However, IR can be of some help in the case that the boundary between morphology and syntax is not clear-cut. Several syntactic changes may result from morphological changes which in turn result from phonological changes. The thesis

presupposes that the application of IR to syntactic reconstruction can be successful in the case that the syntactic change occurs as a result of the phonological or morphological change.

The grammaticalization process as described in §3.2.2.1 can be used as a technique to hypothesize the earlier stages of CNNCs. This can be clearly seen in the cases where one of the variant forms contains a grammaticalized constituent. For example, if a language has the constructions {N,NUM} and {N,NSG} where the non-singular marker, typically dual and trial markers are grammaticalized from the numerals *two* and *three* respectively, by means of grammaticalization, it can be hypothesized that {N,NUM} is the earlier form. This fits well with Givon's (1971: 413 in Hopper and Traugott 2003: 26) well-known maxim, "Today's morphology is yesterday's syntax". In this case, IR can be applied.

Turning to CM, this method is used to reconstruct a given linguistic feature at the level of the language family by means of comparing languages in the family. CM is in fact a method which is primarily used for examining a genetic relationship within a language group. It can be used to reconstruct a given linguistic form when the genetic relationship of the language group has been established (Harrison 2003: 225). In CM, the historical linguist exploits alternations taken from different languages within a family as evidence to reconstruct the proto-form in the proto-language. Therefore, the method of CM is based on the assumption that the alternating forms have descended from a single source in the proto-language.

Like IR, in CM, among those related forms, some forms may represent the original form, whereas the others represent later stages in the historical pathways. For example, in the Jukun languages (Niger-Congo; Nigeria), namely H₂ne, Jibə, Wapa, Djyi, Waphə and Wapan, there is a variety of word forms for the lexeme 'person'. Such a variety of forms is used as a hypothetical evolutionary ladder as shown in Figure 3.1 below. The order of the emergence of forms is explained by generalizations about phonological change.

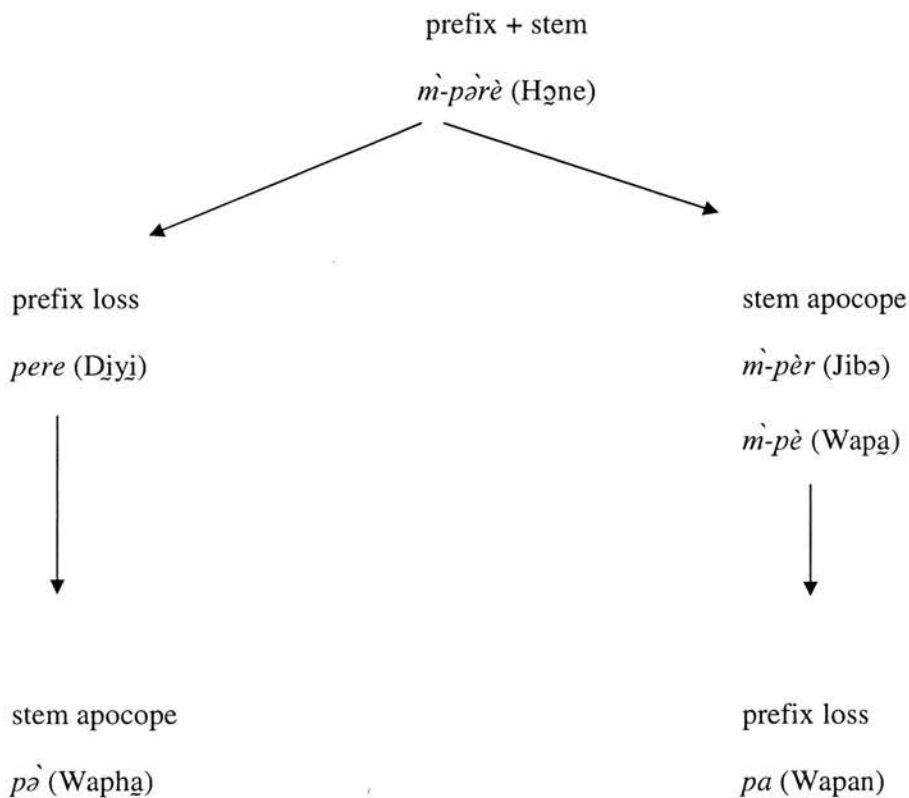


Fig. 3.1 Morphological reconstruction of the lexical form 'person' in Jukun (Storch 1997: 169)

The figure shows the various forms of the lexeme 'person' which are attested in the modern Jukun languages. These forms mirror various stages of historical development of the word forms since the earliest stage (i.e. *m̀-ɸàr̀è*) till the latest stage (i.e. *ɸà* and *ɸa*). From Figure 3.1, some languages appear to be undergoing a change that is now complete in other languages. Hɔ̀ne preserves the original word structure for 'person' in Jukun. In Djɪɪ, the prefix *m̀-* has been lost. In Jibə and Wapə, the prefix *m̀-* remains but the final syllable of the stem is dropped. In Waphə and Wapan, the prefix and final syllable have been lost. Overall, this is an example of the application of CM exploiting the variety of the present forms to investigate the diachronic processes of change. The various forms shown are related by the rules of sound change.

It is generally known that CM is successful with phonology and morphology but less successful with syntax (Fox 1995: 104). Even so, a few principles can be regarded as useful but limited in use, such as typologically-based reconstruction. As the principle has already been mentioned in IR, this principle will not be described again here. The difference is just that in IR, the linguist looks at the alternating forms only in one language. On the other hand, in CM, the linguist compares the alternating forms in languages in the family. For example, if the languages in the family generally have number-marking systems, it would be conjectured that the proto-language is supposed to have a number-marking system. If any language lacks the number marking system, the language is conjectured to have lost the number-marking system.

In addition, there is another principle which is implemented by CM in syntactic reconstruction and can be helpful for the current project if used with caution. This method uses the *majority situation*. We firstly need to establish the predominant patterns of daughter languages in the family. For example, consider the position of the verb in the following languages:

<i>Sanskrit</i>	<i>final</i>
<i>Hittite</i>	<i>final, with initial as a marked variant</i>
<i>Latin</i>	<i>final, with initial as a marked variant</i>
<i>Greek</i>	<i>final, initial, or second position (enclitic)</i>
<i>Old Irish</i>	<i>initial</i>

In this case, based on the majority situation in the daughter languages, verb final position should be presumed to be the unmarked pattern, with verb initial position as a stylistically marked variant for Proto-Indo-European (Fox 1995: 106-107).

This principle seems to be simple, but as Fox pointed out, there are a few problematic issues we need to be aware of. The first is that when the number of alternating patterns distributed in the family is not significantly different, there is no principle for determining which case deserves priority. Also, the feature under study is perhaps no more or less influenced by the interference of neighbouring languages through language contact in the area. Thus, the feature may not originally belong to the

proto-language (Fox 1995: 108). Moreover, the principle is workable only when the classification is valid and the data collected represent all or almost all the branches of the family. Once the hypothesized structural pattern in the proto-language is identified, the current patterns in the daughter languages can be presumed to develop from the reconstructed pattern in the course of time.

3.2.2.4 Language contact

The last theoretical framework for an investigation of the historical development of CNNCs concerns *language contact*. Generally speaking, *language contact* refers to a linguistic situation where the system of one language (the source language) affects another (the receiving language) via a certain interaction between the speakers of the two languages (including through any kind of media such as texts or television). Language contact will result in linguistic variation and perhaps a change in the receiving language (Thomason 2001: 1-3; 2006: 340). The change caused by language contact is generally known as *contact-induced change*. Given that many cases in our sample show that structural patterns of CNNCs undergo change after contacting a language with different types of CNNCs, this section briefly reviews certain aspects of contact-induced change, focusing on structural borrowing.

According to Thomason (2006: 340), there are several types of contact-induced change other than the obvious case of borrowing foreign material. For example, for the case of slow language death, the linguistic material in the receiving language may be lost, but the dying language and the replacing language do not turn out to be more similar. Arabana (Pama-Nyugan; South Australia) is an example of such a language. In this language, the trial marking is being lost after contact with English, but the plural marking as used in English is not required in CNNC_{NSG} (cf. Hercus 1994: 64-65). Also, some changes may happen to linguistic structures of the receiving language as an indirect result from a borrowing of morphemes. A good example of a phenomenon such as this is provided by numeral classifier languages such as Japanese in which the numeral classifier constructions presumably developed highly after the borrowing of numerals from Chinese (see §8.3.1.2 for more details). Moreover, a contact-induced

change can be motivated by a mixture of internal pattern pressures and contact factors. In Finnish (Uralic; Finland), for example, presumably under the influence of Baltic languages, the structural pattern of CNNCs has become similar to those of Baltic languages in a certain aspect, but at the same time Finnish maintains its own unique characteristic (see §9.5.2 for more details). These types of language-induced change may be observed in the current data.

Language contact concerning CNNCs involves structural borrowing. It has often been mentioned that while lexical borrowing is fairly common, the direct borrowing of structures is rather limited. Structural borrowing may be possible only when the grammatical systems of the contact languages are similar to each other (Meillet 1921:87 in Thomason 2006: 339) or when the foreign features fit the tendencies of development in the receiving language (Jakobson 1962: 241 in Thomason 2006: 339). However, Thomason and Kaufman (1988: 15) strongly claim that linguistic structural constraints can in fact be overcome by social factors such as “relative population sizes, length of contact and degree of bilingualism” (Thomason and Kaufman 1988: 65-66). The classic situation for such structural borrowing may take place in a speech community with a subordinate population speaking a genetically unrelated language to a numerically dominant population. The subordinate population later becomes bilingual. Over generations, those bilinguals’ descendants will speak the borrowing language with some aspects of the dominant population’s grammar. For example, a number of grammatical features in Turkish have been borrowed into Asia Minor Greek, or from Indic to Northern Dravidian (Thomason and Kaufman 1988: 67).

According to Thomason and Kaufmann (1988: 74-76), both lexical and structural borrowing depends on the intensity of contact, which can be divided into 5 borrowing scales as follows:

(1) *Casual contact*: At this stage, the non-basic vocabulary, especially words denoting cultural and technological items are borrowed due to cultural pressures, for example, the loanwords in English like *ballet* from French or *spaghetti* from Italian.

(2) *Slightly more intense contact*: At this stage, function words such as conjunctions are borrowed.

(3) *More intense contact*: At this stage, more function words are borrowed such as derivational affixes abstracted from borrowed words, including *low numerals*.

(4) *Strong cultural pressure*: At this stage, there can be a transmission of major structural features that cause little typological change.

(5) *Very strong cultural pressure*: At this stage, major structural features causing typological disruption interfere with the language, for example, a change from a prefixing language into a suffixing language.

To conclude, structural borrowing is quite possible, especially when the languages are in intense contact. In fact, the claim can be evidenced by areal syntactic features that can be observed in all major linguistic areas (see Aikhenvald and Dixon (2001) for comprehensive studies on areal diffusion in major linguistic areas). This review of language contact gives us some background knowledge about how structural borrowing may occur. This is useful for the investigation of the historical development of CNNCs which are subject to change due to language contact.

3.3 Postulating the evolutionary trajectories of CNNCs

The method used in postulating the evolutionary trajectories of CNNCs is to use a variety of current forms of CNNCs in modern languages as a hypothetical evolutionary ladder. The earliest stage before the rise of CNNCs is called *Stage 1*, and the later stages developed from there are called *Stage 2*, *Stage 3* and so on. Based on the results obtained from the historical examination in §3.2, we will have information about historical origins of each structural type. The next step is to tie up these structural types into a network which can be simply depicted as follows.

Supposing there are three types from three languages, namely type A from L(language)1, type B from L2 and type C from L3. These languages need not be related to each other in any sense (i.e. genetically, typologically or geographically). Based on the results obtained from the historical investigation on the origins of each type in §3.2, in L1, type A developed from type B. However, the historical data available in L1 are not old enough to the original source for type B. So, to examine a history of type B, we need to look into L2, in which this type is currently used. Based on the historical investigation in L2, type B develops from type C which belongs to L3. However, supposing further that type C is the earliest stage, the procedure to reconstruct the scenario then stops at type C. Therefore, at this stage the evolutionary scenario of historical development of type A, B and C can be displayed graphically as in Figure 3.2. The arrow symbol ↓ represents ‘becomes’.

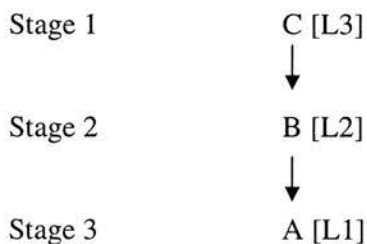


Fig. 3.2 Basis of reconstruction

Figure 3.2 shows that type C changed into type B, and type B in turn changed into type A. The languages in the square brackets represent the languages where the changes occur. For example, L1 represents a language in which the change from type B into type A is observed. Note that this does not mean that L1 has changed from L2. It is emphasized that the three types shown in the evolutionary trajectory are currently attested in the modern languages. In fact, the trajectory can be depicted as in Figure 3.3 below:

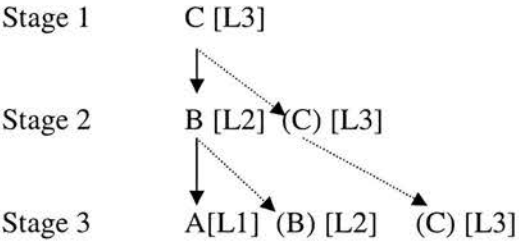


Fig. 3.3 Basis of reconstruction: co-existent types

Figure 3.3 illustrates the co-existent types. For example, at Stage 1, type C changed into type B, which appears at Stage 2, but type C still exists in the world’s languages as evidenced by L3. At Stage 3 which is the most current stage, all the three types are present in modern languages. However, to simplify the hypothetical evolutionary trajectories of CNNCs (see Chapter 10), the information about co-existent types (i.e. types in parentheses with dotted arrows) is not displayed in the trajectories.

In some cases, one type can be a source for two or more types. For example, type B can be a source for type A and type D. In this case, the trajectory can be presented as in Figure 3.3 below.

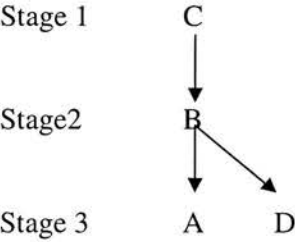


Fig. 3.4 Basis of reconstruction: two types derived from the same source

In addition, type A in some languages may change from type B but in other languages type A may change from type C. In this case, type A may be present at different stages as shown in Figure 3.5, where type A is displayed at Stage 2 and Stage 3.

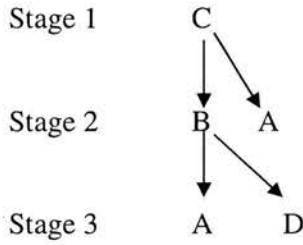


Fig. 3.5 Basis of reconstruction: same types derived from different sources

It is possible that there might be more than one origin of historical paths of CNNCs, and hence more than one trajectory for the evolutionary scenario of $CNNC_{SG}$ and $CNNC_{NSG}$. In this case, the evolutionary trajectories can be displayed as in Figure 3.6.

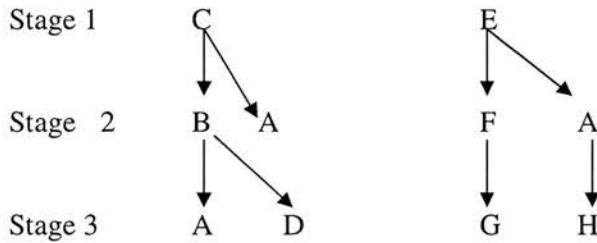


Fig. 3.6 Basis of reconstruction: two different origins of the current types

Figure 3.6 illustrates two origins of the current types A, B, C, D, E, F, G, and H. These types have different origins. At Stage 1, there are two types, namely C and E which are not related to each other and do not develop from any other types. Each type changed into other types, creating their own evolutionary path.

The procedure employed in the current approach is quite similar to IR and CM in historical linguistics. That is to say, the three methods use a variety of forms which are supposed to descend from a single form to reconstruct a historical development of the forms under study.

However, IR and CM are different from the current method used here in an important respect. That is to say, those methods assume that the various forms in question are *genetically related*, whereas the current method is not based on that assumption. The languages under consideration are not confined to genetic relationship, and the method is hence labeled as *Intergenetic Comparative Method*, in parallel with *Intergenetic Grammaticalization Comparison* used in Heine and Kuteva (2002b and 2007, to be discussed below later). This is because we are dealing with the evolution of language at a *global level*, with the time depth dated back to the first appearance of numerals in human language, say approximately 5,000-10,000 years.⁴ When time-depths exceed 5,000 years, the classical historical methods may not be so useful. As is generally recognized, CM is most effective at a local level (at the level of language family) and not beyond 5,000 years. But when looking at the evolutionary paths of certain constructions at a *global level* with longer time depth, we need a comparison which can be applied across families, assuming that some families may preserve the earlier characteristics.

The approach adopted here in fact is rather similar to the approach used in Heine and Kuteva (2002b and 2007) in drawing the evolutionary trajectory of grammatical forms in human language. Their approach is called *Intergenetic Grammaticalization Comparison*, which means “comparison [of grammaticalization] across the boundaries of language families (or phyla)” (Heine and Kuteva 2002b: 377). In other words, it is a comparison that ignores the genetic relationship between languages under consideration. The procedure is described as follows.

⁴ It is noted in Sidwell (1999: 254) that from the evidence of linguistic reconstruction and the simplest contemporary counting system in the most primitive societies, the counting system is possibly a recent development taking place around the late Neolithic (i.e. New Stone Age, dated back to approximately 10,000-5,000 BC), the transitional stage from foraging to farming economy. If indeed this hypothesis is right, the earliest CNNCs in human languages should be dated back not later than 10,000 BC.

- a. X and Y are phenomena that are related in some way
 - b. Hypothesis 1: X existed prior to Y.
 - c. Hypothesis 2: There was a change X > Y (but X continues to exist parallel to Y).
 - d. There is evidence in support of (c)
 - e. There are specific factors that explain (c)
- (Heine and Kuteva 2007: 20)

With this procedure, Heine and Kuteva (2007: 111) have established some salient pathways of grammaticalization. These pathways are taken together and have been conflated into an evolutionary trajectory of grammatical development. The trajectory is reproduced as in Figure 3.7 below:

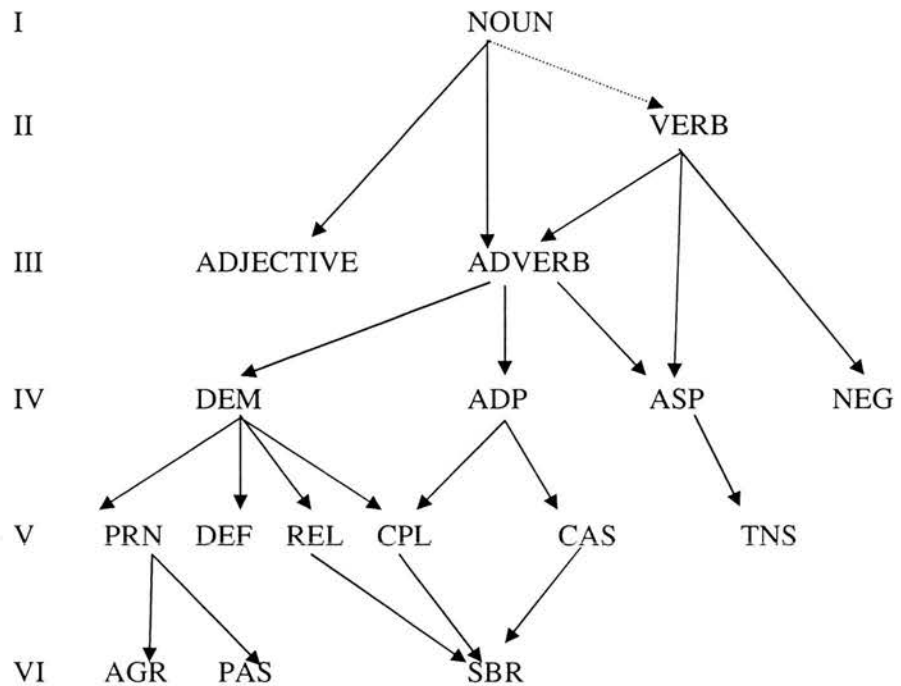


Fig. 3.7 Layers of grammatical development

Abbreviations: I, II, etc. = layers; AGR = agreement marker; ADP = adposition; ASP = (verbal) aspect; CAS = case marker; CPL = complementizer; DEF = marker of definiteness (“definite article”); DEM = demonstrative; NEG = negation marker; PASS = passive; PRN = pronoun; REL = relative clause marker; SBR = subordinating marker of adverbial clauses; TNS = tense marker. (Heine and Kuteva 2007: 111)

It is noted in Heine and Kuteva that their work is not entirely innovative but rather similar to previous research dealing with the early human language (e.g. Sankoff 1979; Comrie 1992; Aitchison 1996 in Heine and Kuteva 2007: 24). The common assumption of these works is that “certain kinds of present linguistic alternation can be reconstructed back to earlier states without that alternation” (Comrie 1992, 2002 in Heine and Kuteva 2007: 24).

Although the procedure and assumption employed in the current project are not different from those in Heine and Kuteva’s (2002b, 2007), the term *evolution* employed in this project is not exactly the same as that used by them. They refer to:

[R]egularities in the development of linguistic forms and structures based on crosslinguistic observations. The development from a numeral ‘one’ to an indefinite article, for example, is an instance of an evolution since it can be observed to occur regularly and independently across languages. (Heine and Kuteva 2002b: 378)

The difference lies in the issue of *regularities*, implying that several instances can be observed cross-linguistically. In the current project, historical evidence is scarce, so some historical pathways (for example, a change from type A to type B) may be evidenced by only one or two languages. Also, unlike grammaticalization, in which the diachronic change seems universal (i.e. the pattern of change can be found across time and place), some diachronic pathways of CNNCs are probably idiosyncratic. In some cases, there might be only one language illustrating that pathway. In other words, the nature of development of grammatical categories and CNNCs is not really the same. For practical reasons, this project will not be confined to the regularities.

3.4 Conclusion

This chapter has presented the means employed in the current project to find out what the evolutionary trajectories of CNNCs are like. The means by which the thesis investigates types and the historical origins of types of CNNCs appear to be basic approaches used in general in typological and historical linguistics. These approaches are not strikingly innovative or theoretically deep. However, the method of *Intergenic Grammaticalization Comparison* employed in Heine and Kuteva (2002b and 2007) to examine the very distant past of human language is relatively new and practical for historical investigation at a greater time depth (compared to CM). The current thesis also uses the similar method labeled as *Intergenic Comparative Method* with a sizeable typological database. Although the method used in the thesis looks similar, the difference is that the method is not confined to regularities.

4 Before the Rise of Cardinal Numeral-Noun Constructions

Before embarking on a description of the diversity of CNNCs in modern languages, one might be tempted to question what the linguistic scenario prior to the rise of CNNCs in early languages might have been. As pointed out earlier, the presence of a *cardinal numeral* is the defining component of CNNCs. For this reason, it might simply be argued that any linguistic quantificational expressions without cardinal numerals, or *non-numeral quantificational expressions* (henceforth *NNQEs*), used in contemporary languages can mirror the scenario which may have existed prior to the emergence of CNNCs. This argument is based on two assumptions. The first is that the commonness of certain linguistic features across languages which are genetically, geographically and typologically unrelated may suggest that they are archaic features which were present in early language. This assumption is made for cases other than those of lexical borrowing, such as English *igloo*, due to language contact through modern media. Secondly, the general characteristics of languages, as well as the nature of the changes observed in languages spoken today, are the same as those of the past. This assumption arises from the principle of *uniformitarianism*. The reader is referred to §11.3.9 or to Heine and Kuteva (2007: 28-32) for more details.

The term *NNQEs* here is used in the sense of linguistic expressions only. Non-linguistic quantificational expressions, such as counting by means of *one-to-one correspondence* (i.e. using materials such as sticks or pebbles to match with the things being counted) is disregarded here, because they are considered as extra-linguistic quantificational usage.

In this chapter, four types of *NNQEs* observed in human languages are discussed as follows, suggesting a diachronic hypothesis about the emergence of the phenomena. It is suggested that at the earlier stages, humans might have used *reduplication*, which is a process of doubling the words to distinguish number meaning. This proposal is

discussed in §4.1. The other simplest way is to use a lexical device, namely non-numeral quantifiers for number distinction (§4.2). A somewhat more advanced stage is to use grammatical means of number distinction (§4.3). The first three types involve imprecise number, that is, the number distinction made is only between *one* and *more than one*. The final type of quantification is the use of words with a numerical interpretation, which has the potential to develop into numerals in later stages of language history, leading to the emergence of CNNCs accordingly (§4.4). Finally, the rise of CNNCs is touched on briefly in the final section (§4.5).

4.1 Reduplication

Reduplication refers to a morphological process in which the entire word or a part of it is repeated for semantic or grammatical purposes. Repetition of a word which does not show altered lexical or grammatical meaning, but just a strong emphasis, such as the English expression “...*no no no (please)*” does not count as reduplication (Inkelas 2006: 417; Rubino 2005: 114; Bakker and Parkvall 2002: 1). Examples of reduplication can be found in Dyirbal (Pama-Nyungan; Australia) and Wambaya (West Barkly; Australia), as shown below.

(4.1) *Dyirbal (Dixon 1972: 242)*

ɟalŋga ɟalŋga

girl girl

'girls'

(4.2) *Wambaya (Nordlinger 1998:43, 106)*

bung-mungmaji

REDUP-old.man

'old men'

Example (4.1) illustrates *total reduplication* (alias *full reduplication*) due to the fact that the entire word (i.e. *ɟalŋga*) is doubled. On the other hand, example (4.2) illustrates *partial reduplication* because only a portion of the word (i.e. *bung* in *bungmaji*) is repeated.

Reduplication serves several functions. For example, reduplication may be used to denote plurality on nouns, as exemplified in (4.1) and (4.2). For verbs, reduplication can be used to express the continuation or repetition of an action as in Yidjŋ (Dixon 1977: 96) e.g. *fafama-n* ‘jump’ versus *ʃafafafama-n* ‘jump a lot’. In addition, as for adjectives and adverbs, reduplication may be used to express intensity, as in Thai *dam* ‘black’ versus *dám dam* ‘very black’. Reduplication can carry a number of other meanings (see, for instance, Rubino 2005: 114-115), but here we consider only

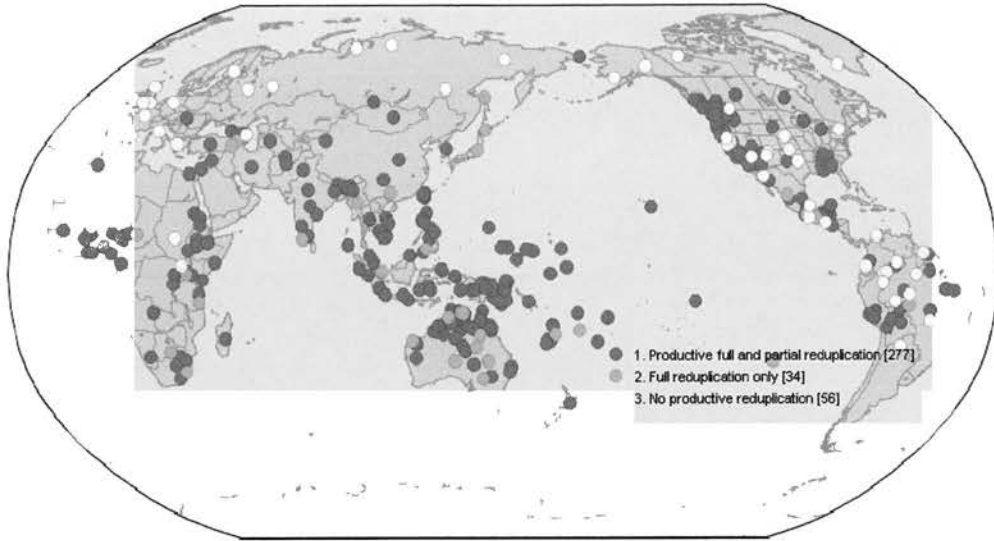
reduplication with a pluralizing function on nouns (henceforth *plural reduplication*) as being relevant to the issue of NNQEs.

In examples (4.1) and (4.2), it can be observed that the reduplication clearly reflects an iconic similarity between the forms and their meaning in terms of quantity. That is, the repetition of form suggests the greater number of referents. Although the word form appears twice however, whether totally or partially, it does not necessarily mean that the meaning of the reduplicated form is equivalent to *two*. Rather, the meaning is equivalent to *more than one*. The correspondence of linguistic form and its meaning as shown in these examples is generally known as *iconicity*. Plural reduplication therefore exemplifies the principle referred to in Haspelmath (to appear: 1) as “the iconicity of quantity”—that is, “greater quantities in meaning are expressed by greater quantities of form”.⁵

Givón (1995) explains iconicity in terms of biology. He provides biological evidence to argue that the roots of iconic coding can be found at the cognitive and neurological levels. More detailed discussion on the biological basis of iconic codes can be found in Givón (1995: 59-65). If iconicity is indeed a matter of neurological structure, the argument that plural reduplication exists at the pre-cardinal numeral-noun constructional stage is better supported.

In terms of frequency and distribution of reduplication, the use of reduplication is a fairly common phenomenon in the world’s languages. In Rubino’s (2005: 114-117) sample of 367 languages across the globe, 311 languages are found to have reduplication—that is, about 85 per cent of the sampled languages. According to Rubino’s map, reproduced as Map 4.1 below, reduplication is likely to be productive in most parts of the world, except Europe, where reduplication is relatively rare.

⁵ Although reduplication involves iconicity, not every meaning carried by reduplication is iconic, especially in instances of partial reduplication (Inkelas and Zoll 2005: 14). For example, in Illocano (Austronesian; Philippines) the reduplication applied to numbers specifies *limitation*, such as *wal-waló* ‘only eight’ (Rubino 2005: 115). This function does not reflect an iconic nature in the reduplication process.



Map 4.1 Geographical distribution of reduplication (Rubino 2005: 116-117)

Although reduplication seems very common, unfortunately the survey does not indicate how many of the languages which have it use reduplication as a method of pluralization. Dryer (2005) surveys various grammatical devices which languages employ to indicate plurality on nouns. It is noted that only 8 out of 957 languages indicate plurality by using *full* reduplication as a *primary* method of pluralization. The number of languages with plural reduplication is unexpectedly small, and all are from Oceania. This is probably because partial reduplication is categorized as plural affixes instead of as reduplication proper, and reduplication as a non-primary method of marking pluralization is not considered at all. If reduplication as a non-primary method of plural indication is also considered, the number of languages with plural reduplication is likely to be much higher than reported in Dryer's survey. However, because there is so far no literature which focuses intensively on the frequency and distribution of plural reduplication cross-linguistically, we do not know exactly how frequently plural reduplication occurs in the world's languages.

However, Kajitani (2005) has explored 16 languages representing genetic and geographic diversity to test Uspensky's (1972: 70 in Kajitani 2005: 93-94) hypothesis about the preferential ranking of the four semantic properties of reduplication. They include *augmentation* (increase of quantity) of which plural reduplication is an instance, *diminution* (decrease of quantity; e.g. Agta, spoken in the Philippines, *wer* 'creek' versus *walawer* 'small creek'), *intensification* (increase of degree; e.g. Turkish *dolu* 'full' versus *dopdolu* 'quite full'), and *attenuation* (decrease of degree; e.g. Swahili *maji* 'wet' versus *maji-maji* 'somewhat wet'). The reader is referred to Kajitani (2005: 98-99) for more examples and references therein. Kajitani summarizes the result in a diagram reproduced as Figure 4.1.

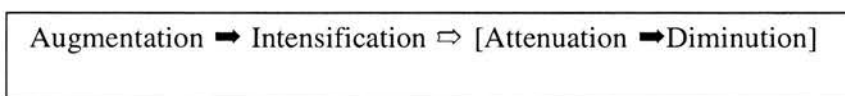


Fig. 4.1 Hierarchy of semantic properties expressed by reduplication (NB: $X \Rightarrow Y$ stands for 'X is universally preferred over Y in every language,' and \Leftrightarrow stands for statistical preference" (Kajitani 2005: 94))

The diagram suggests that *augmentation* is universally preferred over the other three properties (i.e. it holds in all instances). The second preference is *intensification*, which is statistically preferred over *attenuation* and *diminution*. Finally, *attenuation* is universally preferred over *diminution*. In other words, the presence of decrease (diminution and/or attenuation) implies the presence of increase (augmentation and/or intensification). As pointed out by Kajitani, although this is a small data set and the generalizations remain tentative, the results reflect the iconic nature of the reduplication process and common perceptual experiences of speakers across languages.

At this stage, when considering the generalizations made by Kajitani along with the widespread occurrence of reduplication in the world's languages, we might predict that plural reduplication should be fairly common. This is because *augmentation* is ranked first in the hierarchy. Yet, further typological research with a larger sample on plural reduplication is required to confirm Kajitani's argument.

Because of its iconic nature, it is quite possible that plural reduplication may be frequently found. If indeed plural reduplication is common cross-linguistically, it is reasonable to predict further that plural reduplication might have been used as a grammatical device to express plural number in early language.

4.2 Non-numeral quantifiers

Non-numeral quantifiers here refer to the words or phrases used to indicate an imprecise amount of things, such as English *many*, *a few*, and *some*. Examples from a couple of languages are given below. The reader is referred to Appendix 3 (Question 7) for more examples of non-numeral quantifiers.

(4.3) *Tok Pisin (Creoles; Papua New Guinea; Verhaar 1995: 31)*

plantì man
many man
'many people'

(4.4) *Acoma (Keresan; US (New Mexico); Maring 1967: 119)*

hác²Ū sái
several day
'several days'

(4.5) *ǀAni (Khoisan; Botswana; Heine 1999: 39)*

thû xèù
many hippo
'many hippos'

The use of non-numeral quantifiers is probably the simplest and most universal way to indicate the plurality of referents in human languages, even in languages with the simplest systems of quantification, such as Nadëb (Vaupés-Japurà; Brazil; Weir 1984: 103-4 in Epps 2004: abstract), which has only the numeral *one* and then *a few* and *many*. An exception to this may be Pirahã (Mura; Brazil), which is claimed not to have terms for quantification such as *all*, *some*, and the like. It is noted, however, that there are words that can be loosely translated as 'many', namely *bá a gi so* (lit. 'cause to come together') and *ogí* 'big' (Everett 2005: 622-623). Since the use of non-numeral quantifiers seems to be universal, it is reasonable to argue that the use of lexical devices such as these might have been present in the early stage of quantificational expressions.

4.3 Grammatical means of number distinction

A number of languages employ grammatical means to mark the opposition between singular and plural. These are generally known as number markers. One of the widespread grammatical methods is the use of third person plural pronouns (i.e. *they*) to code nominal plurality. Heine and Kuteva (2002a: 237-238) have demonstrated that the plural markers can be created from third person plural pronouns by a grammaticalization process. Baka (Niger-Congo; Cameroon) provides helpful evidence. More examples from a variety of languages can be found in Heine and Kuteva (2002a: 237-238).

(4.6) *Baka* (Christa Kilian-Hatz, personal communication, in Heine and Kuteva 2002: 237)

(a) *wósé* *wó* *á* *gɔ*
 woman *3PL* *ASP* *go*
 ‘The women are going.’

(b) *wósé-o* (*wó*) *á* *gɔ*
 woman-3PL *3PL* *ASP* *go*
 ‘The women are going.’

Based on grammaticalization theory, variant (a) should be an early construction, whereas variant (b) should be a developed form, as shown in the phonetic erosion (*wó* > *o*) when the third person plural pronoun grammaticalizes into a plural marker. It is not mentioned whether or not the two variant forms are different in terms of pragmatic and stylistic meanings. The use of the third person plural pronoun as a plural marker is common in Creoles. An example is Krio (English Based Creole; Sierra Leone)

(4.7) *Krio* (Todd 1979: 288 cited in Heine and Kuteva 2002a: 238)

(a) *dɛm* *bin* *futam*
 3PL *TNS* *shot*
 ‘He/She/It was shot (by them).’

(b) *mi* *padi* *dɛm* *buk*
 1SG.POSS *friend* *3PL* *book*
 ‘*my friends’ book*’

According to Dryer (2005a), there are other grammatical means to express number. These include: plural affixes as in English *dogs* where *-s* is a plural marker; changes within the noun stem as in English *man* (singular) versus *men* (plural); plural tone (i.e. a tone for forming plurality on nouns) such as Ngiti (Kutsch Lojenga 1994: 135 in Dryer (2005a: 138) *kamà* ‘chief’ *kámá* ‘chiefs’; plural words (i.e. “words whose meaning is like that of plural affixes, but they are separate words” (Dryer 2005a: 139); and plural clitics (a morpheme expressing plurality attached *phonologically* to whatever words in the noun phrase). The former (plural words) is illustrated by Chalcatongo Mixtec (Oto-Manguan; Mexico) in (4.8), and the latter (plural clitics) by Sinaugoro (Austronesian; Papua New Guinea) in (4.9).

(4.8) *Chalcatongo Mixtec (Macaulay 1996: 113 in Dryer 2005a:139)*

ni-xáǎ=rí *kwaʔà* *žúʔa* *káni* *xináʔa*
 COMP-buy=1SG *many* *rope* *long* *PL*
 ‘*I bought many long ropes*’

(4.9) *Sinaugoro (Kolia 1975: 124 in Dryer 2005a: 139)*

belema *bara=ria* *taulatoitoi*
 python *big=PL* *six*
 ‘*six big pythons*’

Relying on the simple logic that the number marking should not be required when the number of the noun was made clear by the presence of numerals, it is conjectured that in the languages where these grammatical means of plurality are used, these grammatical means might have been present prior to the rise of numerals.

4.4 Words with a numerical interpretation

The rise and expansion of numerals in a language are motivated by economic pressure as well as non-economic factors like technological development (Winter 1999). In some contemporary *materially primitive societies* (i.e. communities of foragers, hunter-gatherers, herders or farmers who live in small-scale local communities and use simple technologies, Kuper 2004: 800-803), numerals play almost no significant role in people's daily life. These societies may have numerals just for *one* or *two*, as observed in some Australian aborigines (Dixon 2002:67), or they may perhaps even lack native real numerals at all, as evidenced in some South American indigenous languages (Closs 1986). However, the people in those societies appear to use a word or phrase which is basically non-quantificational but has a “numerical interpretation” (Hurford 1999: 12).

It is pointed out in Everett (2005, for example) that “Pirahã [(Mura; Brazil)] is the only language known without number, numerals, or a concept of counting.” (2005: 622). The claim has remained controversial, especially if it is interpreted that Pirahã *totally* lacks numerals, and that Pirahã is the *only* language lacking numerals (see Nevins, Pesetsky and Rodrigues 2007: 36-40, for counter-arguments). However, relying on the examples below, it seems likely that Pirahã *at least* has a word with a numerical interpretation because the word *hói* ‘small size or amount’ as shown in (4.8) below can be interpreted as the numeral ‘one’ in *some contexts* (Everett 2005: 623).

(4.10) Pirahã (Everett 2005: 623)

a.	<i>tí</i>	<i>'ítí'isi</i>	<i>hói</i>	<i>hií</i>	<i>'aba'átgio</i>	<i>'oogabagáí</i>
	<i>ISG</i>	<i>fish</i>	<i>small</i>	<i>PRED</i>	<i>only</i>	<i>want</i>
	<i>'I only want {one/a couple/a small} fish.'</i>					

NB: “This could not be used to express a desire for one fish that was very large, except as a joke”. (NB original)

b.	<i>tiobáhai</i>	<i>hói</i>	<i>hií</i>
	<i>child</i>	<i>small</i>	<i>PRED</i>
	<i>'small child/child is small/one child'</i>		

In this case, Pirahã may not be so different from some other South American Indian languages. The striking characteristic is only that the word with a numerical interpretation is perhaps highly ambiguous, and that the meaning of the word relies on context to a great extent.

In Chiquito (Chiquito; Bolivia), there was no real word for number. The people used the word *etama* ‘alone’ to express the number concept of *one* (Conant 1896: 2). In Carib (Cariban; Suriname), the word for *two* has the root meaning ‘break, split’ (Menninger 1969: 119). In Hup and Yuhup (Guaviare-Japura; Brazil and Columbia; Epps 2004: abstract), the numeral *one* is both similar to the demonstratives ‘that’ and ‘other’. In Wari’ (Chapacura-Wanhan; Brazil), the native real numerals are not available, however, so the verbs *xica’ pe* ‘to be alone’ for *one* and *tucu caracan* ‘to face each other’ (rarely used nowadays) for *two* are used (Everett and Kern 1997: 347-348).

(4.11) Wari’ (Everett and Kern 1997: 348)

- | | | | | |
|----|---|-------------------|-----------------|----------------|
| a. | <i>xica’</i> | <i>pe</i> | <i>na</i> | <i>tarama’</i> |
| | <i>alone</i> | <i>be.at.SBJ</i> | <i>3SG.RP/P</i> | <i>man</i> |
| | <i>‘There is one man.’ (lit. ‘The man is alone.’)</i> | | | |
| | | | | |
| b. | <i>‘tucu-u</i> | <i>caracan</i> | <i>na</i> | <i>xirim</i> |
| | <i>face-1SBJ</i> | <i>each.other</i> | <i>3SG.RP/P</i> | <i>house</i> |
| | <i>‘There are two houses.’ (lit. ‘The houses face each other.’)</i> | | | |

In addition, the body-part counting system (i.e. the system expressing particular numbers by pointing to the body parts such as fingers, hands, wrists, shoulders, eyes) which is widely observed in Papua New Guinea (Lean 1985-6 in Comrie 1999: 82) may also imply the previous stage right before the rise of CNNCs. This is the numeral system used in Haruai (Upper Yuat; Papua New Guinea; Comrie 1999: 81-94). In this system, the numeral *one* is related to the word denoting ‘little finger’. Comrie (1999: 83) notes that,

[...]despite the transparent relation to tallying by using body-parts—as seen in the possibility of identifying a numeral by pointing to the appropriate body-part without uttering its name—the body-part expressions used as numerals can be integrated into fluent Haruai speech, usually following the head noun (as is also usual for adjectives in Haruai). There is thus no reason to suppose that the expressions of the body-part system are in any sense ‘extra-linguistic’.

This statement implies that the body-part system in (modern) Haruai is not entirely extra-linguistic—in other words, the system is not purely based on the use of body parts without being accompanied by body-part terms. This claim is justified by the evidence that body-part terms are used with quantified nouns—that is, the body-part term functions as a numeral. However, the interesting implication is that at the earlier stages, body parts might have been used for tallying things or expressing the number of things, perhaps with or without words. The difference in numerical values of the quantified nouns may have been indicated by the different body parts, and so by the different body-part terms accordingly. This means that the body-part terms were also used to serve a quantifying function which is the secondary function, but along with a pointing gesture. Later, when these body-part terms became commonly known in the speech community, they were then used as numerals proper, as evidenced by their combination with quantified nouns. The process finally led to the rise of CNNCs in the language. Probably, this also happened in some other languages with body-part counting systems.

Based on the assumption of uniformitarianism, these examples taken from languages spoken in primitive societies could reflect the state of affairs which may have characterized the pre-CNNCs-stage. That is to say, during the period when humans did not have numerals, they made use of words with a numerical interpretation, i.e. internal resources in their language which could express low numbers like *one* and *two*. These words would often mean ‘alone’ or ‘small’ or ‘little finger’ for *one*; and ‘things in pairs’ for *two*. Based on etymological work on the historical origins of numerals in some languages (such as Sanskrit *éka*- ‘one’ or ‘alone’ (Burrow 2001: 258)), it might be predicted that if these primitive societies did not borrow numerals from other languages, these words with a numerical interpretation might eventually develop into numerals.

Due to the fact that numerals are not totally grammatical lexemes (see §3.2.2.1), and that the sources for numerals in some cases are in fact perhaps more grammatical than the numerals themselves (for example, in Hup the numeral *one* is associated with the demonstrative *that*), the term *grammaticalization* may not really match the process. Instead, let us label the particular processes of grammatical-semantic change in numerals as *numeralization*⁶ where a linguistic item develops a numeral concept. Nevertheless, in some (or perhaps most) cases, numeralization can count as grammaticalization when the numerals in the language are clearly grammaticalized from nouns or verbs and currently behave as adjectives or quantifiers. According to grammaticalization theory, the adjectives are more grammatical than nouns or verbs (Heine and Kuteva 2002b: 383).

A major mechanism involved in the numeralization process is semantic extension—gaining a numeral concept, especially by metaphorical extension (i.e. transferring a concept from one word to another based on conceptual similarity, for example, *alone* and *one*; *wings* and *two*). The mechanism of semantic bleaching (i.e. losing the original lexical meaning) may also be involved in the process where the words with a numerical interpretation have gradually lost their original meaning over time. An example is Indo-European **tr*⁷ for the number ‘three’ (e.g. English *three*, French *trois*), which is related to the root **ter* ‘beyond’ (Luján Martínez 1999: 207). Due to a decrease in semantic transparency, some of the proposals of numeral reconstruction so far are not convincing and do not achieve consensus.

To summarize, words with a numerical interpretation and numerals which are semantically transparent or semantically somewhat opaque (with known historical

⁶ Readers may have come across the term *numeralization* with different senses. For example, in Senft (1996: 19-20, and personal communication), the term (*quantifying*) *numeralization* refers to a situation where a classifier not only classifies nouns, but also expresses the quantity of the nouns. An example is taken from Kilivila (Austronesian; Papua New Guinea)

<i>ma-kupo-na</i>	<i>yena</i>
<i>this-two.string-this</i>	<i>fish</i>
<i>‘these two strings of fish’</i> (Senft 1996: 19)	

The classifier *kupo* not only classifies the quantified fish as a ‘string’, it also inherently expresses the number of strings, i.e. ‘two (strings)’ (cf. English ‘a pair of shoes’).

⁷ Examples preceded by an asterisk are historically reconstructed forms.

origins) are empirical evidence which is suggestive of the resources that humans in a pre-numerical stage used to express precise number. The differences between numerals in terms of semantic transparency may reflect the chronological continuum of the emergence of numerals, where the more semantically transparent the numerals are, the more recently they are likely to have developed.

4.5 The rise of numerals and the polygenesis of CNNCs

As illustrated in §2.1, CNNCs are basically composed of a noun and a numeral. In accordance with this definition, the emergence of CNNCs is conditioned by the fact that the lexicon of a language must contain these two kinds of linguistic item, namely nouns for naming things and numerals for naming numbers. According to linguistic evidence (Hurford 1987 and 2001)⁸ and neurological studies on infant arithmetic, a human's brain can distinguish small numbers like 1, 2, and 3 easily without counting. In fact, even animals and pre-linguistic humans have primitive mathematical concepts (Simon, Hespos and Rochat 1995; Dehaene 1997). However, this does not mean that every language is expected to have numerals. Pirahã (Mura; Brazil; Everett 2004 and 2005) in which numerals are claimed to be absent exemplifies an exception. This is because, as mentioned earlier, the rise of numerals depends both on economic pressures and various cultural factors (Winter 1999). The use of numerals is also triggered by being in contact with neighboring languages which have numerals. For example, Wari' (Chapacura-Wanhan; Brazil) has used the Portuguese numerals since its contact with European/Brazilian society (Everett and Kern 1997: 347-348).

It is argued in Wiese (2007) that in early language, humans would have had a limitation in numerical thinking. That is, numerical thinking in humans relies on finite and iconic representations as can be observed in animals and human infants. The emergence of low-valued numerals and counting sequences would contribute to the development of numerical thinking in humans, and this in turn would open the way to building up more numerals infinitely. Wiese therefore refers to the phenomenon in which the expansion of numerals and numerical thinking support each other as “the co-evolution of number concepts and counting words” (Wiese 2007: 1).

⁸ Cross-linguistically, the grammatical idiosyncrasies of numerals up to about 4 are reflected in grammatical systems such as number systems (singular/dual/trial); word classes (i.e. adjective-like) and suppletive forms of ordinals (e.g. English *one/two/three* versus *first/second/third*). The phenomena concerning the idiosyncrasies of the small numerals suggest that the low-valued numerals might have emerged earlier than the high-valued numerals due to the human ability to subitize these low-valued numerals (Hurford 1987 and 2001).

Obviously, numerals in the world's languages arose in different times and places—that is, by *polygenesis*. Whereas the class of nouns is likely to be commonly observed in languages, the class of numerals is not. The emergence of numerals in a language is therefore regarded as a crucial factor which contributes to the rise of CNNCs in that language, since a numeral is the defining characteristic of a CNNC. Because the rise of numerals is by polygenesis, and because the emergence of CNNCs is conditioned by the rise of numerals, the emergence of CNNCs is accordingly also by polygenesis.

4.6 Conclusion

In this chapter, it has been conjectured that before the rise of CNNCs in early language, humans might have employed NNQEs, the quantificational expressions which can be observed in contemporary languages. Four means of non-numeral quantification have been proposed, namely reduplication, non-numeral quantifiers, grammatical means of number distinction and words with a numerical interpretation. Due to its iconic nature, reduplication is probably the simplest means of number distinction. Non-numeral quantifiers and grammatical means of number distinction involve lexical devices and grammatical devices respectively. The two means of NNQEs probably lead to a common type of $CNNC_{NSG}$, namely $\{N, NUM, NSG\}$ depending on the obligatoriness of the non-singular marker on nouns in the language (see Chapter 9 for further discussion). Finally, the use of words with a numerical interpretation is likely to be the most important, due to the fact that they are a potential source for numerals and hence the emergence of CNNCs. Since the emergence of numerals is linked to cultural development, which varies from society to society, the emergence of numerals and CNNCs obviously happens by polygenesis. Studying the scenario which existed before the rise of CNNCs (such as the rise of numerals and non-singular elements which are constituents of CNNCs) helps to understand the emergence of CNNCs themselves.

5 On the Extra Elements

When collected on structures of CNNCs, one will find that they may include other elements (or *extra elements*) apart from a noun and a numeral. These extra elements can be divided into two groups. One is relevant to quantification, whereas the other is not. Let us label the former as *quantificational extra elements* and the latter as *non-quantificational extra elements*. Only the quantificational extra elements play a part in classifying structural types of CNNCs. The current study is therefore concerned with extra elements of this kind. These quantificational extra elements contribute to the diversity and complexity of structural types of CNNCs, as will be illustrated in the following chapters. There are several kinds of quantificational extra elements which are required in CNNCs, typically number markers and numeral classifiers. Although the quantificational extra elements may seem redundant when they appear in CNNCs (since they do not add any new information to the constructions), they are nonetheless relevant to quantification in certain ways, and hence to CNNCs.

This chapter is devoted to providing operational definitions and outlining the historical development of the quantificational extra elements employed in CNNCs so as to aid our understanding of how they are relevant to quantification and CNNCs. The quantificational extra elements involved are number markers (§5.1), numeral classifiers (§5.2), and other less prevalent elements (§5.3). Non-quantificational extra elements include, for instance, gender markers and copulas, which can be regarded as present in CNNCs simply due to the general requirements of the structural rules of the particular language. Since these extra elements are not relevant to quantification, they are not taken into account in classifying structural types of CNNCs, although we do touch on them briefly (§5.4).

5.1 Number markers

Number markers refer to a grammatical device of making meaning distinctions with regard to number in word classes. This is typically done by means of personal pronouns and nouns, but as this thesis deals with cardinal numeral-noun constructions, the number-marking systems considered here will be concerned with *nouns* only. The issue of number-marking systems plays a significant role in the thesis, as the complexity of CNNCs partly involves number-marking systems. This section first describes characteristics of languages with a number-marking system and then provides a short overview of the historical development of number markers, namely the *singular* (and *singulative*), the *plural*, the *dual*, the *trial*, the *non-plural* and the *inverse number*, though with particular concentration on the singular and the plural.

5.1.1 Number-marking systems

A language with a number-marking system in nouns is a language where a means of expressing a difference in the number of referents can be found within the grammatical system. For example, in English, in *These peaches are ripe*, there are three signals which indicate that there is more than one peach, namely the demonstrative adjective *These* as opposed to *This*, the *-es* plural suffix, and the subject-verb agreement *are* as opposed to *is* (Cruse 1994: 2857). Among these three signals, the *-es* plural suffix is an instance of a number marker on the noun itself, and so it exemplifies the expression of number through *morphology*. On the other hand, the demonstrative adjective *these* and the verb *are* represent the expression of number through *syntax*, as the number markers are shown somewhere else in the sentence (Corbett 2000:136-138). Therefore, English represents a language which has both morphological and syntactic number-marking systems.

Another example of a language with a number-marking system is Maung (Australian; Northern Territory). Maung exemplifies a language where number marking is not on the noun itself but on the article preceding the noun. Examples are given below.

(5.1) *Maung* (Capell 1970: 50)

- a. *dja* *arargbi*
 ART.SG *man*
 'the man'
- b. *bada* *arargbi*
 ART.PL *man*
 'the men'

Note that when indicating singularity of the noun *arargbi* 'man', the article *dja* is used, but for plurality the article *bada* is used. *Maung* can be regarded as a language with a number-marking system because a means for expressing number on nouns can be found within the grammatical system—that is, through use of a system of articles.

A language without a number-marking system in nouns is one where the means for making distinctions in number cannot be found within the grammatical system. The noun in this type of languages is vague with regard to number—that is, the noun can be interpreted either as expressing *one* or *more than one* unless the noun is accompanied by quantifiers. According to a widely held view (see §5.2.4), such nouns have the semantic property referred to as *transnumerality* (Bisang 1999), a term which corresponds to *general number* in Corbett (2000: 10). Languages of this type may use non-numeral quantifiers only when the number distinction is emphasized. These quantifiers are basically lexical items meaning *a few*, *some*, *several*, *many* and the like, as shown in (5.2).

(5.2) *Thai* (own knowledge)

- a. *khǎw* *sǎ̌ǎ̌* *nǎ̌ŋsǎ̌ǎ̌*
 3SG *buy* *book*
 'He buys a book.' or 'He buys books.'
- b. *khǎw* *sǎ̌ǎ̌* *nǎ̌ŋsǎ̌ǎ̌* *lǎ̌ay* *lēm*
 3SG *buy* *book* *many* CLF (lit. 'volume')
 'He buys many books'

In practice, to deem whether a particular language has a system of number-marking is often not a straightforward matter. This is because the marking of number may not apply to all nouns in the language, but rather only to some nouns, typically those denoting humans (Haspelmath 2005: 142). Such a variation in number marking can be explained in the light of the *animacy hierarchy*—i.e. a generalization that nouns or their equivalents are treated differently in language depending on their *degree of animacy* (i.e. a degree to which nouns are regarded as sentient). In general, nouns referring to things such as humans (and especially personal pronouns) are ranked very high in the hierarchy. These nouns are likely to be marked or treated specially in language. On the other hand, inanimate nouns or abstract nouns which have lower animacy tend to be unmarked for number (see, for example, Corbett 2000: 54-94 for a detailed discussion on the animacy hierarchy and grammatical number).

In addition, the marking of number involves *obligatoriness*. That is, in some languages, the marking of number is obligatory, but in others the marking is optional. Relying on the dimensions of animacy and obligatoriness, Haspelmath (2005: 142) divides languages into 6 types with regard to the occurrence of nominal plurality, namely (1) *no nominal plural* (2) *only human nouns, optional* (3) *only human nouns, obligatory* (4) *all nouns, always optional* (5) *all nouns, optional in inanimates* and (6) *all nouns, always obligatory*. The last type is the most common.

5.1.2 *Singular and singulative*

The term *Singular form* refers to a form which denotes one item. In most languages, including English, the singular form is unmarked—that is, there is no linguistic device encoding the singular number (e.g. *man*). However, some languages employ a grammatical device to express singularity. This device is generally known as a *singular marker* or *singulative marker*. Very often the term *singular marker* overlaps with the term *singulative marker* in terms of what they refer to. However, the term *singular form* seems slightly different from the term *singulative* in the following respect. Generally, the term *singular* is used to contrast with the term *plural* (or *dual* or *trial*) in the singular/plural (/dual/trial) number systems. Although cross-linguistically

the singular form is usually unmarked, there exist languages which employ a singular marker, for example, Mohawk *ska-hańselah* [SG-lamp,light] ‘one lamp, light’ (Bonvillain 1973: 235). On the other hand, the term *singulative* is used to contrast with the term *collective* form (i.e. formally singular, but semantically plural) or *general* form (i.e. *general number*, as in example (5.2a) above) from which the singulative form is usually derived (Corbett 2000: 17). For example, Burushaski (isolate; Pakistan) *sisʌn* ‘a person’ and *gərkən* ‘a pea’ are derived from the general noun *sis* ‘person, people’ and the collective or plural noun *gərk* ‘peas’ respectively (Lorimer 1935: 48). The singulative form is therefore conceptualized as the lexical form (typically the general form or collective form) attached with a singulative marker (Corbett 2000: 17). Overall, the singular marker and singulative marker may overlap in terms of semantics, but they are different in terms of their origins and grammatical functions.

Both the singular and singulative markers count as extra elements required in CNNCs because they are relevant to quantification. To see this, we may consider their origins. In the case of the singular marker, according to the current data, one of the possibilities is that the singular marker is likely to be grammaticalized from the numeral one, or related to the numeral one. For example, in Pame (Oto-Manguean; Mexico) the petrified singular prefix *n-* (e.g. *n-ɛi^o* ‘tooth’), which is frozen and serves no grammatical function might have been derived from the numeral *nada* ‘one’ (Manrique Castanade 1967: 346); and in Mohawk (Iroquoian; New York, Quebec), the singular marker *ska-* and the numeral *v’hska* or *énska* ‘one’ originated from the verb root *-t* ‘be one’ (Marianne Mithun, personal communication). The reason why the singular marker is claimed to be relevant to quantification is that historically it was used for number distinction—that is, as the numeral *one*.

Turning to the singulative marker, like the singular marker, a good deal of evidence shows that the singulative marker is derived from the numeral *one*. Again, Burushaski provides a clear example. In this language, the singulative forms are derived from collective nouns by being suffixed with *-ʌn* or *-ən*. The suffixes are probably

derived from the numeral *hAn* ‘one’ (Lorimer 1935: 46). The singulative markers can be regarded as relevant to quantification. As mentioned in the preceding paragraphs, we can see that a language which has the singulative marker such as Burushaski has so-called *plural nouns* or *collective nouns*. When the collective nouns are counted, the nouns are required to be individuated first. A grammatical device of individuation is the singulative marker which is grammaticalized from the numeral one as noted in Lorimer (1935: 46-48):

The suffix *-An*, *-En*, which is probably to be identified with the *-An* of *hAn* one, is added in general to singular forms of nouns and noun-equivalents[...] It has the force of a singular suffix when used with a noun or pronoun, of which the form is not definitely plural, but which is habitually, or at least frequently, used with plural force[...]

The use of singulative markers for nouns which are culturally perceived as a group or mass is common in the world’s languages. These nouns tend to denote masses consisting of isolable members. To refer to an instance of the group, the singulative markers are therefore required. This can be seen in Modern Breton, a Celtic language spoken in France (Press 1986: 70), which has the forms *per* ‘pears’ vs *per-enn* ‘a pear’; and also in Arabic (Zabba 2002: 4), in the forms *naml* ‘ants’ vs *naml-at* ‘an ant’. In Tucanoan languages spoken in Brazil and Columbia (Barnes 1999: 221), the base forms of some nouns denoting insects such as gnats and bees are plural; the nouns form the singular by attaching the singulative suffix *-bĩ* or *-wĩ*. Unfortunately, the historical origins of these singulative markers are not mentioned in the sources. It is possible that the singulative markers may derive from some nouns in addition to the numeral *one*, typically the classifier-like nouns, such as the word meaning ‘unit’ or ‘piece’.

It seems likely that the function of the singulative marker is to individualize the noun referent. This function may be similar to that of *numeral classifiers* in classifier languages (cf. §5.2). Whereas numeral classifiers can co-occur with all numerals however, the singulative form is unlikely to be present in the context of numerals greater

than one. However, in Arabic and perhaps some other languages, the singulative forms can be present in dual or plural constructions, as shown below (Zabbal 2002: 118).

(5.4) *Arabic* (Zabbal 2002: 118)

a. *al-baqar-at-aa-ni*

*DEF-cow.COLL-SG-DU.NOM-ni*⁹

'the two cows'

b. *al-baqar-Ø-aat-u*

DEF-cow.COLL-SG-PL.FEM.NOM

'the cows'


5.1.3 Plural

Plural markers refer to a grammatical device for indicating more than one referent, or, in the case of languages where dual or trial number systems exist, more than two or more than three referents respectively. As was already mentioned in Chapter 4, there are various ways for languages to show number distinction without numerals (e.g. reduplication). These ways were referred to as *non-numeral quantifiers*. This suggests that nouns might have occurred with non-numeral quantifiers during the time that numerals did not exist in a language. Lehmann (1995: 57) points out that there are two stages of a possible origin of nominal number: first, the number marker (typically the pronoun) accompanies the noun only when plurality is emphasized, and then later, this marker becomes affixal and increasingly obligatory. In what follows, building on Lehmann's (1995: 56-59) idea of the grammaticalization of number along with the principle that diversity across languages reflects the evolutionary stages of current forms (cf. §3.3), the thesis proposes 6 *degrees (or stages) of obligatoriness* in the marking of plural number, in the continuum shown below. This organization of degrees of obligatoriness mirrors a *hypothetical* evolutionary ladder of plural marking. For example, the structures in the 3rd degree are conjectured to develop from the 2nd degree,

⁹ The suffix *-ni* is not glossed in the original source, however, according to Clive Holes (personal communication), *-ni* is part of the dual suffix *-aani* in the nominative.

which in turn might have developed from the 1st degree. Historical evidence from languages is required to support the idea of change from one stage to another.

Non-obligatory

- 
- 1) None (number expressed through context, e.g. through use of 'many')
 - 2) Optional, used only when emphasized
 - 3) Obligatory without numerals but absent with numerals
 - 4) Obligatory without numerals but optional with numerals
 - 5) Obligatory either with or without numerals
 - 6) Bound to the stem

Obligatory

Fig. 5.1 Degrees of obligatoriness of number marking (paralleled with grammaticality of number marking—from lexical form to bound grammatical form)

Figure 5.1 shows 6 degrees of obligatoriness of number marking in human languages, from no number marking at all to absolute obligatoriness (i.e. when the number marker is an obligatory part of the word). It is not necessary that one language display only one degree. Some languages may show various degrees of obligatoriness because they have several alternate strategies of number distinction.

1st Degree: No number marking (number is known from contexts such as 'many')

To start with, languages do not use a grammatical number on nouns at all. The nouns in these languages are vague in terms of number. The number distinction of the noun may be known by the context, especially from the use of non-numeral quantifiers meaning 'many' or 'all' and determiners meaning 'these', for example. This characteristic is observed in the majority of East and Southeast Asian languages, and also in most pidgins and creoles.

Some of these non-numeral quantifiers are lexico-grammatical (i.e. having a grammatical function but remaining semantically transparent) rather than being absolutely grammatical. These non-numeral quantifiers are elements that may

potentially be grammaticalized into plural markers. Heine and Kuteva (2002a: 36) provide examples from several languages showing that plural markers are derived from the quantifier *all*. For example, French *tout les* ‘all the’ is grammaticalized into the nominal plural *proclitic* (i.e. a clitic placed before the word to which it is phonologically attached) or into the prefix *tule, tle, te* in Tayo Creole French as shown in (5.5).

(5.5) *Tayo Creole French* (Kihm 1995: 234, 237 in Heine and Kuteva 2002a: 36)

<i>tle</i>	<i>fler-la,</i>	<i>le</i>	<i>fini</i>	<i>puse</i>	<i>e</i>	<i>pi</i>	<i>sa</i>
PL	flower-DEF	TAM	CPL	grow	and	then	3PL

<i>atra-de</i>	<i>puse</i>	<i>akor</i>
PROG	grow	still

‘The flowers have been growing, and they are still growing.’

2nd Degree: Optional, used only when emphasized

At this degree of non-obligatoriness, the plural marking-system may exist, but the plural marking is used just when the plural number is specially emphasized. For example, in Chontal Maya (Mayan; Mexico), although plurality can be indicated by the plural suffix *-lop’* (e.g. *čitam-lop’* [pig-PL] ‘pigs’), the plurality on any class of nouns (including humans) is nevertheless not obligatory (Knowles 1984: 202). In other words, even though plural markers exist, they are not normally employed. For example, in Ineseño Chumash (Chumash; California) the noun *ku* ‘person’ has a plural form *kukhu*² ‘people’, but it is not used in a context where the plural sense is presupposed, as shown below (Mithun 1988: 212).

(5.6) *Ineseño Chumash* (Applegate 1972: 458 in Mithun 1988: 212)

<i>s-iy-axi-kum</i>	<i>ha-ku</i>
3PL.ITER-dance	DEF-person

‘The people are dancing.’

3rd Degree: Obligatory without numerals but absent with numerals

As for the third degree, number is marked obligatorily, especially with human nouns or animate nouns, following the principle of the animacy hierarchy. However, the number marker is not used at all if numerals are present. Hungarian is an example of such a language.

(5.7) *Hungarian (Kenesei, Vago and Fenyvesi 1998)*

a. *a* *nyul-ak* *egymás* *után* *el-pusztul-t-ak*
 DEF.ART rabbit-PL each.other after PFX-die-PST-INDEF.3PL
 ‘The rabbits died one after the other.’ (p.272)

b. *sok* *naptár*
 many calendar
 ‘lots of calendars’ (p.230)

c. *három* *fiú*
 three boy
 ‘three boys’ (p.229)

As for this development, it is hypothesized that the use of plural markers with plural referents is increasingly frequent. For this reason, plural marking is becoming systematic, though the use of plural markers is not extended to contexts with quantifiers. This corresponds to the hypothesis that frequency of use may lead to what has been referred to as *entrenchment*, or the establishment of linguistic structures in the speaker’s cognition, as Langacker (1987: 59) puts it:

With repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit; moreover, units are variably entrenched depending on the frequency of their occurrence.

4th Degree: Obligatory without numerals but optional with numerals

For the first three degrees of obligatoriness, the number marker is not supposed to be used if numerals (and perhaps also non-numeral quantifiers) are present. This is

because plurality can be understood from the context. This situation makes sense, as we can see the function of plural marking. Nevertheless, the situation becomes more difficult to understand when nominal plural markers are used in the noun phrase in spite of the presence of numerals. In some languages, the presence of grammatical numbers in CNNCs does not apply to all nouns. This may suggest that the number system in the language is increasing or declining. However, this does not affect all noun classes at the same time. The tendency is for the development of the number system to be associated with human nouns first and then for it to spread to animate and inanimate nouns (following the principle of the animacy hierarchy). The reverse may be the case when there is declining use of a number system.

This degree may correspond to a language with a mixture of {N,NUM} and {N,NUM,NSG} (see Chapter 6). It also reflects the fact that the use of a plural marker may not spread to the whole nominal system, unlike English and other Indo-European languages where most countable nouns are marked for plural. Rather, plural marking is restricted to only human or animate nouns at this stage. This can be illustrated in Jarawara (Arawá; Brazil) where animate nouns are marked for plural, whereas (masculine) inanimate nouns are not, as shown below. Note that the morpheme *mee* is an augmentative modifier as well as the 3NSG pronoun marking the non-singular number (Dixon 2004: 302).

(5.8) Jarawara (Dixon 2004)

a. {N,NUM}

<i>jobes</i>	<i>joro</i>	<i>ni-kime</i>
<i>house.M</i>	<i>stand</i>	<i>AUX-Two.M</i>

'two houses stand (there).' (p.152)

b. {N,NUM,NSG}

<i>hijama_o</i>	<i>mee</i>	<i>otaa</i>	<i>tao</i>
<i>peccary</i>	<i>3PL</i>	<i>IEX.A</i>	<i>shoot</i>

<i>ka-ni-kima-mina</i>	<i>otaa-ke</i>
<i>APPL-AUX -two-morning.F</i>	<i>IEXC-DECL.F</i>

'We shot two white-lipped peccaries this morning' (p.155)

5th Degree: Obligatory either with or without numerals

At this degree, plural markers are always employed either with or without the presence of numerals. Plural markers are clearly redundant in contexts with quantifiers including numerals. It is conjectured here that due to the frequent use of plural markers in the language, the number distinction becomes the nominal category that the speaker is aware of. It might be possible that in the speaker's mind the unmarked noun or the absence of a plural marker implies the singular meaning instead of general number. Consequently, when the speaker refers to the plurality of the referents, the form of the noun in the speaker's mind can no longer be a singular form. Rather, the form would be subconsciously changed into a plural form corresponding to the meaning that the speaker wants to express. The characteristic of nouns in the 5th degree may correspond to Rijkhoff's (2002) "singular object nouns", namely, situations where in the speaker's mind the noun has "a definite outline in the spatial dimension" (Rijkhoff 2002: 53). As already mentioned in §2.2.2, a singular object noun has a spatially bounded and non-agglomerative property. According to Rijkhoff, this property is associated with the compulsory plural marking on nouns of plurality. The absence of plural markers implies the meaning of singular number (not general number). Thus in general, when the noun conveys the meaning of plurality, the use of the singular form is ungrammatical and is in fact conceptually impossible. English and other Indo-European languages are examples of languages at this degree.

Regarding development, once the plural marking became entrenched in the speaker's cognition as described in the 4th degree, this plural marking might have been extended to contexts with numerals as well, when numerals were introduced in these languages. This implies that numerals came into existence after the entrenchment took place. In any case, it seems likely that the intensive and frequent use of plural markers on countable nouns may change the type of nouns from general nouns to singular object nouns. This phenomenon would involve three interactive domains, namely language use, language acquisition and language change. The framework dealing with these issues is referred to by cognitive linguists as the *usage-based model*, but it will not be pursued

here. For further discussion on usage-based models, the reader is referred to Croft and Cruse (2004: 291-327) among others.

6th Degree: Bound to the stem

As for the highest degree or the last stage of the hypothetical evolution of plural marking, number marking is morphologically obligatory because the number markers have changed into an obligatory part of the words, presumably through the processes of *grammaticalization* (cf. §3.2.2.1) and *lexicalization* (cf. §3.2.2.2). At the earlier stage, the grammatical number in nouns in these languages might have been expressed by affixes (which are grammaticalized from the lexical or less grammatical forms such as third personal plural pronouns) attached to the noun stem (i.e. the part of the word which is not yet inflected). Over time, the affixes have been fused into the stem, and have decreased in compositionality and have hence turned into a lexicalized item through the process of lexicalization. The stem of nouns is basically bound; that is, it cannot stand alone without an affix encoding the noun class along with the singular and plural. The Bantu languages spoken in the central and southern part of Africa are best known for this degree of obligatoriness. In Bantu languages, generally, singular/plural alternations are marked with affixes attached to a noun stem. The noun stem cannot be used independently. For example, the noun stem *-ntu* in (5.9a) and (5.9b) cannot be used independently without the noun class prefixes *mu-* and *a-*.

(5.9) *Lunda* (Kawasha 2003: 124)

a. *mu-ntu* *wu-mú*
I-person *I-one*
'one person' (Class I for human nouns and singular)

b. *a-ntu* *a-yedí*
II-person *II-two*
'two people' (Class II for human nouns and plural)

There is no historical evidence available for the development of the plural prefix *a-* in Lunda. The noun classes in Niger-Congo languages are fully grammaticalized and so it is extremely difficult to trace back their original meanings (Williamson and Blench 2000: 13). However, it is noted in some places (e.g. Williamson and Blench 2000: 13; Heine and Reh 1984: 234) that the number-class markers on human nouns are associated with the third person plural pronouns.

In many African languages, 3rd person plural pronouns have been added to nouns and have developed into nominal plural markers...In accordance with this strategy, the personal pronoun follows the specified unit (Heine and Reh 1984: 234).

Baka (Niger-Congo; Cameroon), which was already shown in §4.3, provides a clear example.

(5.10) *Baka* (Christa Kilian-Hatz, p.c. cited in Heine and Kuteva 2002: 237)

a. *wósé* *wó* *á* *gɔ*
 woman *3PL* *ASP* *go*
 ‘The women are going.’

b. *wósé-o* (*wó*) *á* *gɔ*
 woman-PL *3PL* *ASP* *go*
 ‘The women are going.’

The example in (5.10a) illustrates the use of *wó* ‘third person plural pronoun’ as a separate plural marker which remains semantically transparent and at the same time functions as a pronoun. On the other hand, in (5.10b), *wó* (which has changed to *-o* due to phonetic erosion) has assumed a new grammatical function, i.e. the full plural suffix, via the grammaticalization process. This is often presumed to be a source for the affix denoting person-gender-number (PGN) in Niger-Congo languages. This is not surprising due to the inherent meaning and function of the third person plural pronoun (Frajzyngier 1997: 209).

5.1.4 Other number markers

5.1.4.1 Dual marker

Dual markers are grammatical forms which basically express the number *two* of their referents, and perhaps also have a special pragmatic meaning as shown in the examples (5.11b and 5.12b) below. The dual marker can be a core element when it is used instead of the numeral *two*, but it is regarded as an extra element when it is accompanied by the numeral *two*. It is generally acknowledged that the dual comes historically from the numeral *two* through the process of grammaticalization. Kayardild (Tangkic; Queensland) and Ngiyambaa (Pama-Nyungan; New South Wales) provide clear instances.

(5.11) *Kayardild* (Evans 1995: 184)

a. *ngijin-da kiyarrng-ka kularrin-da*
ISG.POSS-NOM two-NOM sister-NOM
'my two sisters'

b. *ngijin-da kularrin-jiyarrng-ka*
ISG.POSS-NOM sister-DU-NOM
'my two sisters'¹⁰

(5.12) *Ngiyambaa* (Donaldson 1980: 102)

a. *bulagar miri*
two dog
'two dogs'

b. *miri-bula:*
dog-DU
'a pair of dogs'¹¹

¹⁰ The difference in meaning of the two examples remains obscure. It is probably the case that the dual would implicate joint action, while the numeral would not, but this has not been confirmed by a native speaker (Nick Evans, personal communication).

¹¹ It is noted in Donaldson (1980: 102) that "The dual suffix *-bula:* means 'in a group of two.' " Therefore, it seems the dual suffix *-bula:* and the numeral *bulagar* differ slightly in meaning. That is, the numeral is used to convey the number of referents, while the suffix not only conveys the number of referents but also implies that the referents are in the same group (of two).

The dual suffixes in (b) are obviously related to the numeral ‘two’ in (a) because they are similar in form and meaning and can be explained in terms of the grammaticalization process. However, they are morpho-syntactically different. That is, the numeral ‘two’ in (a) is a lexical form or content word occurring independently, while the dual marker in (b) must be attached to the noun. According to grammaticalization theory, the duals in Kayardild and Ngiyambaa are likely to originate from the numeral *two*. There are two interrelated mechanisms involved in the grammaticalization process, namely decategorization and erosion. The numeral *two* which has developed into the dual has lost its lexical status. It has changed from a lexical form into a grammatical form used for marking the dual number. Therefore, it undergoes decategorization. Also, it can be observed that in Kayardild the consonant *k* in *kiyarrng* ‘two’ has changed to a semi-vowel *j* in *jiyarrng*, making it easier to be suffixed to other morphemes (or resulting from being always attached to other morphemes). In Ngiyambaa, the final syllable *gar* in *bulagar* ‘two’ has been lost due to certain phonological processes. The phonetic change as such is referred to as *phonetic decay* or *phonetic erosion*. More evidence can be found in Dixon (2002: 117) and Heine and Kuteva (2002a: 302-303).

5.1.4.2 Trial marker

The term *Trial marker* refers to a grammatical marker basically denoting three entities. In some languages trial may refer to three or a little more than that (Corbett 2000: 21). Similar to the dual marker, the trial marker can be a core element when it is used *as* the numeral *three*, but it is regarded as an extra element when it appears *along with* the numeral *three*. It is generally acknowledged that the trial originates from the numeral *three* through the process of grammaticalization. Ambrym (Austronesian; Vanuatu) provides examples of the trial marked on personal pronouns and other word classes. The trial marker is grammaticalized from the numeral *sol* ‘three’ (Heine and Kuteva 2002a: 297).

(5.13) *Ambrym* (Paton 1971: 16, 24, 44-46 in Heine and Kuteva 2002a: 297)

- a. *vɛɛn* *ŋa-sul*
 woman *PCL-TRI*
 ‘three women’
- b. *gam-sul*
 2PL-TRI
 ‘you three’

5.1.4.3 Non-plural marker

The term *Non-plural marker* refers to a grammatical marker denoting that the noun to which the affix is attached is not plural. This marker is observed in Imonda (Border; Papua New Guinea). In this language there are a few nouns having an inherent plural number which refers to the number greater than ‘two’. So, when the nouns are modified by the numerals ‘one’ and ‘two’, the non-plural (NPL) must be used as shown in (5.14). Otherwise, the meaning of the numerals and the inherent number of nouns are not compatible.

(5.14) *Imonda* (Seiler 1985)

- a. *toad-ianèi* *mugasl*
 boy.PL-NPL *one*
 ‘one boy’ (p. 62)
- b. *agō* *-ianèi* *sabla*
 woman.PL *NPL* *two*
 ‘two women’ (p.39)

5.1.4.4 Inverse number marker

The term *Inverse number marker* refers to a grammatical number marker used for changing the basic number of nouns. This kind of number marker is found in Jemez (Kiowa-Tanoan; US, New Mexico) and is the unusual grammatical feature shared by the

Kiowa-Tanoan languages.¹² In this family, nouns are categorized into classes based on their basic number, i.e. the inherent number of the noun. For example, in Jemez, the basic number of Class I nouns is singular. When the suffix *-š* is added to the noun of Class I, the suffix will indicate dual and plural. The basic number of Class II nouns is plural. When the inverse suffix is added, the noun will be changed to singular and dual (Yumitani 1998:97). This applies both in contexts with numerals and without. Examples (5.15a) and (5.15b) illustrate the use of the inverse number suffixes without numerals and with numerals respectively.

(5.15) *Jemez (Yumitani 1998)*

- a. *pæš ší* 'A deer (basic-SG) fell off.'
pæšš ijí 'Deer (INV-DU) fell off.'
pæšš eš í 'Deer (INV-PL) fell off.' (p.100)

- b. *wî mš-sá-š* *ɬil-kʷâ*
two cat-INV TR-put.down/PFV
 'I put down two cats.' (The basic number of *mš-sá* 'cat' is singular.) (p.150)

¹² The inverse number is a relatively rare grammatical phenomenon in human languages, however, the grammatical feature as such can also be found in some Oceanic languages (Corbett 2000: 162-166).

5.2 Numeral classifiers

In this thesis numeral classifiers are considered as quantificational extra elements based on the frequently cited view that the genesis of numeral classifiers is attributable to enumeration (e.g. Emeneau 1951: 93 in Allan 1977: 293, Greenberg 1972, Bisang 1999: 113-123). As the complexity of CNNCs is also relevant to this type of extra elements, this section therefore outlines general issues relevant to numeral classifiers. In §5.2.1 and §5.2.2, the characteristics of *numeral classifiers* and *numeral classifier languages* are described. Then, in §5.2.3 an attempt is made to answer the question of how one can distinguish numeral classifiers from other linguistic elements that may appear in the same syntactic slot. Section 5.2.4 looks into the functional motivation of the genesis of numeral classifiers so as to understand how the numeral classifier is relevant to quantification.

5.2.1 Characteristics of numeral classifiers

The *numeral classifier system* is a type of the so-called *classifier systems*. The term *classifier system* refers to “a grammatical system of noun categorization device(s) in a particular language.” (Aikhenvald 2000: vii). The term covers all noun categorization devices cross-linguistically, such as *gender* in European languages, *noun classes* in African and Australian languages and *noun classifiers* in Meso-American Indian and Australian languages. The reader is referred to Aikhenvald (2000) for a comprehensive typological description of these classifier systems.

As for the term *numeral classifier*, like most types of classifiers, it is generally defined as a morpheme categorizing the noun with which it occurs based on semantic characteristics, namely *animacy*, *humanness*, *shape*, and other inherent properties. However, the defining characteristic is that this type of classifier is always required as a unit counter in a quantificational expression in a language which has a numeral classifier system as the dominant mode. For this reason, such a classifier is recognized as a *numeral classifier*, though it is used with non-numeral quantifiers (e.g. *many*) as well.

A numeral classifier may also occur with demonstratives (e.g. *this*) and with adjectives (Allan 1977, Craig 1994: 565 and Aikhenvald 2000: 2). Below is an example of such a classifier construction taken from Thai.

(5.16) *Thai (own knowledge)*
mǎa sǎ:m/lǎ:y tua
dog three/many CLF (lit. 'body')
'three/many dogs'

In (5.16) *tua* is regarded as a numeral classifier, since it categorizes the noun *mǎa* 'dog' as a class of animal. It is always required in the quantificational expression.

Moreover, as noted in many places (for example, Greenberg 1972: 185; Allan 1977; and Dixon 1986: 106) regarding word order in a quantificational expression, the key characteristic distinguishing the numeral classifier system from other classifier systems is that a numeral classifier is likely to form a morphological unit with the numeral, not with the quantified noun. This can be observed from the fact that a numeral classifier normally appears in a position *adjacent to* or *bound to* a numeral. An exception is Kana (Niger-Congo; Nigeria) where a numeral classifier tends to form a morphological unit with the noun instead (Ikoro 1994: 19-23) (see further discussion of this exceptional case in §8.3.2).

According to Lyons (1977: 463) and Aikhenvald (2000: 115), numeral classifiers can be divided into 2 types, namely *sortal classifiers* and *mensural classifiers*.

A sortal classifier is one which individuates whatever it refers to in terms of the *kind of entity* that it is [...] A mensural classifier is one which individuates in terms of *quantity* (Lyons 1977: 463, italics mine).

Below are examples of the two types of numeral classifiers.

(5.17) *Thai (own knowledge)*
a. nók sǎ:m tua
bird three body
'three birds'

- b. *nók* *sǎ:m* *fū:ŋ*
 bird *three* *group (basically used for nouns denoting animals)*
 ‘*three birds*’

The numeral classifier *tua* in (5.17a) illustrates a *sortal classifier* as it suggests the *kind* of the entity (i.e. an animal), whereas the numeral classifier *fū:ŋ* in (5.17b) illustrates the *quantity* of the entity—that is, more than one bird; and categorizing bird as an animal. It is noted in Aikhenvald (2000:115) that

Since the choice of a mensural classifier is often determined by the temporary state of an object (its quantity, or the arrangement it occurs in) there may be more freedom in choosing a mensural classifier than in choosing a sortal one.

For example, in Tzeltal (Mayan; Mexico), the numeral classifier for the noun *lagrio* ‘brick’ is *pech* ‘rectangular, non-flexible object’. However, the noun ‘brick’ can be measured in various ways depending on its arrangements. The numeral classifier for this purpose involves arrangement, such as *latz* for a stack of bricks, *bus* is for a pile of bricks (Aikhenvald 2000: 115). The origin of the mensural classifier perhaps comes from those nouns in a language which can be arranged and measured in various ways.

5.2.2 Numeral classifier languages

A language with a *numeral classifier system* as a dominant mode has productive use of numeral classifiers in quantificational expressions. That is, the use of numeral classifiers is *obligatory* to almost *all* nouns. In these languages, however, some nouns denoting *units*, such as *time* and *distance* (e.g. a noun meaning ‘day’) may not require a numeral classifier as in (5.18), since the noun itself functions like a numeral classifier—that is, as a *unitizer* in the noun phrase (the function of numeral classifiers is discussed in the next few paragraphs). Note also that such a noun is in the same position as the numeral classifiers would be in the language.

(5.18) *Vietnamese (Nguyen 2004: 101)*

bốn ngày

four day

'four days'

In addition, a typical numeral classifier language generally shows a vast number of numeral classifiers, though some languages are exceptions. Some languages use special words similar to numeral classifiers in a quantificational expression, for example, English *bar* and *head* in *three bars of soap* and *two head of cattle*. These words may be referred to as *numeral classifier-like forms* (Downing 1996: 2). Although such numeral classifier-like forms occur with numerals and categorize the nouns in the counting constructions, English should not be referred to as a numeral classifier language. This is because nouns in English normally do not require the numeral classifier in quantificational expressions, and the use of numeral classifier-like forms in the language is too restricted to be referred to as a system.

5.2.3 *Lexico-grammatical items in the numeral classifier slot*

The characteristics mentioned so far separate numeral classifiers from other (lexico-grammatical) forms that may appear in the same syntactic slot, such as *measure words*, *noun classifiers*, *noun class markers*, and *number markers*. These lexico-grammatical forms are described as follows.

5.2.3.1 *Measure words*

Measure words are generally recognized as the lexical words employed to measure a quantity of things. Examples are English *pound* in *two pounds of meat*, *bunch* in *two bunches of flowers* and *glass* in *two glasses of water*. It seems likely that all languages have measure words for measuring mass nouns such as *water* or nouns denoting groups such as *flowers*. In a language with a numeral classifier system, measure words may appear in the same syntactic slot as numeral classifiers, and it seems to be difficult to distinguish between the two types. For example in (5.19), it can be seen that the numeral classifier *lû:k* and the measure word appear in the same slot.

(5.19) Thai (own knowledge)

a.	<i>sôm</i>	<i>sǎ:m</i>	<i>lû:k</i>
	<i>orange</i>	<i>three</i>	<i>CLF</i>
	<i>'three oranges'</i>		
b.	<i>sôm</i>	<i>sǎ:m</i>	<i>laŋ</i>
	<i>orange</i>	<i>three</i>	<i>box</i>
	<i>'three boxes of oranges'</i>		

However, measure words involve the *quantity* and *arrangement* of nouns (Grinevald 2002a: 260), whereas numeral classifiers suggest the *inherent properties* of nouns. Also, numeral classifiers occur with fewer types of nouns, whereas measure words are less restrictive. Moreover, in terms of semantic transparency, the lexical meaning of a numeral classifier is bleached in a quantificational expression, whereas the lexical meaning of the measure word remains unchanged (Aikhenvald 2000: 116-117). As in (5.19a), the noun *lû:k* is regarded as a numeral classifier because it suggests the shape of an orange as a round type of object. Also it is used only with certain nouns, typically those denoting round fruits. The lexical meaning which was originally 'child' dissipates with only the meaning of shape (denoting small and round) remaining in this context. In example (5.19b), the noun *box* is regarded as a measure word because it suggests an arrangement of oranges in a box. Also the noun *box* is used with any nouns that can be *contained* in a box.¹³ The lexical meaning of the word for *box* remains unchanged in this context.

5.2.3.2 Noun classifiers

A *noun classifier system* is a type of classifier system based on the semantic properties of the classified noun. In general, noun classifiers are generic terms meaning *animal, human, vegetable* and *tree*, for instance. The noun *tree* in English *apple tree*, in fact, functions like a noun classifier, (since it indicates that the noun 'apple' being

¹³ Measure words are different from mensural classifiers in that measure words do not classify nouns at all. For example, the measure word *laŋ* 'box' does not classify the quantified noun, since it can be used with any nouns that can be contained in the box. On the other hand, the mensural classifier *fūŋ* 'group' is not used with inanimate nouns, so the use of the mensural *fūŋ* involves the animacy of nouns.

mentioned is a tree, not a fruit). In a given language, not every noun would require a noun classifier.

(5.20) *Yidiny* (Dixon 1982: 185 in Aikhenvald 2006: 465)

<i>mayi</i>	<i>jimirr</i>	<i>bama-al</i>	<i>yaburu-nggu</i>	<i>julaal</i>
<i>vegetable.ABS</i>	<i>yam.ABS</i>	<i>person-ERG</i>	<i>girl-ERG</i>	<i>dig-PST</i>
'The girl dug up the yam' 'lit. 'The person girl dug up the vegetable yam'				

It is noted in Aikhenvald (2000: 84) that

In languages with noun classifiers, distinct classifiers can be used with the same noun to specify the meaning, e.g. Minangkabau *batang limau* (CL¹⁴:TREE lemon) 'lemon-tree', *buah limau* (CL:FRUIT lemon) (Martina 1996 in Aikhenvald 2008: 84).

In other words, noun classifiers are used to clarify the class of nouns which are ambiguous in terms of physical properties in particular languages. As in Minangkabau, the word *limau* may be interpreted as a lemon tree or a lemon fruit, if unaccompanied by the noun classifier *batang* or *buah*.

In the example (5.20) taken from *Yidiny*, it is clear that *jimirr* refers to a class of vegetable, not seeds or something else; *yaburu* refers to a human class, not an animal class, for example. Noun classifiers are also concerned with culture, giving more information about social interaction. For example, in Jalcattec (Craig 1986b: 245) *naj* is a noun classifier for male non-kin, whereas *ho7* is a noun classifier for male kin.

In the case that the noun classifier appears in the same syntactic slot as the numeral classifier, it would be difficult to say whether the morpheme is a numeral classifier or noun classifier. A practical way to approach this is to look at the occurrence of noun classifiers in non-counting constructions such as the example (5.20) above.

¹⁴ In Aikhenvald (2000) CL is abbreviated from classifier.

5.2.3.3 Noun class (or gender) markers

Noun class is a type of a highly grammaticalized classifier system. Unlike numeral classifiers, the noun class assignment to nouns is rather arbitrary. It is alternatively referred to as *gender* (e.g. Corbett 1991, 2005), though traditionally the term *gender* is often reserved for a small noun class system which contains two or three classes of nouns and more semantically opaque (typically masculine and feminine classes), as in French (Aikhenvald 2000: 19). On the other hand, the term *noun class* is often used to refer to the larger noun class systems such as are commonly found in African languages. The systems may contain between five and twenty five classes (Craig 1994: 565). In this thesis, the term *noun class* is generally used to incorporate the gender systems (unless otherwise indicated) because noun class systems always include the gender systems in the traditional sense. In any case, in collecting data, one may question how we can distinguish between noun class markers and numeral classifiers if they appear in the same syntactic position, as they do in Swahili (Niger-Congo; Southern Africa).

(5.21) *Swahili (Welmers 1973: 171)*

<i>ki- kapu</i>	<i>ki-kubwa</i>	<i>ki-moja</i>	<i>ki-lianguka</i>
<i>CL-basket</i>	<i>CL-large</i>	<i>CL-one</i>	<i>CL-fell</i>
<i>'One large basket fell.'</i>			

Although the noun class systems are quite similar to numeral classifiers in that both categorize nouns, there is a significant difference between them. Namely, a noun class system involves elaborate agreement systems (i.e. systems in which a form of elements in the phrase or sentence varies according to the class of the head noun) as in (5.21) where all elements in the sentence agree with the class *ki-* of the head noun *kapu* 'basket', while the agreement in a numeral classifier system is not that complicated, as seen in (5.22) below.

(5.22) *Yurok* (Blevins 2004: 1)

- a. *k.jht.ɬy.ɬy* *pu:k*
 one.CLF *deer*
 ‘one deer’
- b. *kohtoh* *haʔa:g*
 one.CLF *rock*
 ‘one rock’

In Yurok (Algic; California), the numeral agrees with nouns—that is, the numeral co-occurs with the numeral classifier for a round object. The agreement in numeral classifiers is clearly seen in a language where a numeral and a numeral classifier are fused into one morpheme, as in Yurok.

5.2.3.4 Number markers

Mostly, number markers appear next to the noun, such as in English *two dog-s*. However, there exists a language in which a number marker is attached to the numeral. Wolof (Niger-Congo; Gambia and Senegal) is the only language in the current sample illustrating this as shown below.

(5.23) *Wolof* (Ngom 2003: 48)

- n̄aar-i* *xarit*
two-PL *friend*
‘two friends’

In this case, the number marker appears next to the numeral like a numeral classifier. However, the affix *-i* cannot be interpreted as a numeral classifier if the affix is used only with numerals greater than one. So, the numeral classifier and the number marker are different in that the numeral classifier can be used with all grammatical numbers, but the choice of number marker must be compatible with the number of the referents.

5.2.4 A frequently cited view on the genesis of numeral classifiers

An oft-cited typological observation on word orders with regard to the numeral-noun-numeral classifier is that numeral classifiers always appear next to or bound to quantifiers. In other words, the noun is not allowed to be inserted between the numeral classifier and the numeral. Based on this observation, there is a widely held view (e.g. Greenberg 1972; Bisang 1999; Gil 2005) that numeral classifiers are obligatory in some languages for the purpose of counting—that is, “to make count nouns enumerable” (Bisang 1999: 113, and see other functions of the numeral classifiers therein). This is because the typical attachment between a numeral and a numeral classifier suggests that what is counted is not the noun, but the numeral classifier. This also further suggests that count nouns in those languages must have a certain semantic property which makes them uncountable if they are without numeral classifiers.

Therefore, Greenberg looks into non-numeral classifier languages such as English and Arabic to examine the semantic property of nouns which cannot be present in the direct construction of the quantifiers. It is found that those nouns are the so-called *collective nouns* (e.g. English *police, cattle*) and *mass nouns* (e.g. English *water*). The nouns as such require the so-called *unit counters*, namely measure words, numeral classifier-like forms and singulative devices to separate out individual noun referents in a group in the case of collective nouns. They also unitize the non-discrete noun referents in the case of mass nouns. As for the case of *collective nouns*, examples can be found in English and Arabic. In English, the noun *cattle* is a collective noun. It cannot be counted directly. Thus, when the noun *cattle* is counted, the numeral classifier-like form *head* must intervene as in *three head of cattle*. In Classical Arabic, a collective noun cannot be governed directly by a numeral, rather the preposition *min* ‘from’ must be inserted as a singulative device. For example,

(5.24) *Classical Arabic* (Greenberg 1972: 179)

<i>thala:thatu</i>	<i>mina</i>	<i>'ibl</i>
<i>three</i>	<i>from</i>	<i>camel (COLL)</i>
<i>'three camels'</i>		

Turning to the case of mass nouns, one cannot count a mass noun such as *water* directly, for example, because it does not have a discrete boundary. We have to use a noun with a discrete boundary such as *glass*, *cup*, or *bucket* as a unit counter to unitize the referent noun *water* so that it can be counted.

Another interesting finding of Greenberg's about the similarity between a numeral classifier and a quantifier is that what are really counted are not the nouns in the constructions, but rather the individualizers or unit counters. As for the case of collective nouns, Greenberg provides evidence from Omani, an Arabic dialect spoken in Oman as in (5.25).

(5.25) *Omani* (Greenberg 1972: 178)

<i>thala:thit</i>	<i>rwa:s</i>	<i>finda:l</i>
<i>three</i>	<i>CLF.PL</i> (lit. 'heads')	<i>potato</i> (COLL)
<i>'three potatoes'</i> (lit. <i>three heads of potatoes</i>)		

As for the case of mass nouns, English provides good examples. In English, the plural marker appears on the measure words as in *three glasses of water*, suggesting that what is counted is the noun *glass*, not *water*.

Looking at a similar phenomenon where nouns cannot be in a direct construction of the quantifier, Greenberg argues that numeral classifiers presumably have the same function as the unit counters—that is, individualizing or unitizing the quantified nouns. Therefore, nouns in numeral classifier languages are supposed to share a certain semantic property with those nouns requiring unit counters in a quantificational expression. Those nouns are either collective nouns (comparable to English *police*) or mass nouns. Instead of labeling the nouns in numeral classifier languages as collective or mass nouns however, Greenberg refers to them as “transnumeral” (1972: 182), which is

semantically neither singular nor plural. It is a transnumeral category which is neutral in respect to number as opposed to the singulative which involves countability. (Greenberg 1972: 182)

The semantic property termed transnumerality can be illustrated with Thai as shown in (5.26). The noun *nǎŋs ǎ:* ‘book’ which is used without indication of number can be interpreted as *one book* or *more than one book* depending on a given context. Therefore, the noun is regarded as *transnumeral* in Greenberg’s sense.

(5.26) Thai (own knowledge)
khǎw *s ǎ:* *nǎŋs ǎ:*
 3SG buy book
 ‘He buys a book or books.’

Since the bare (count) noun in Thai is not necessarily to be interpreted as a single entity (unlike English), in this sense, the bare (count) noun is not different from the collective noun denoting plurality. Therefore, the noun requires an individualizer when it is counted, as shown below.

(5.27) Thai (own knowledge)
khǎw *s ǎ:* *nǎŋs ǎ:* *sǎ:m/lǎ:y* *lêm*
 3SG buy book three/many CLF (lit. ‘volume’)
 ‘He buys three/many books.’

When the noun is used in a quantificational expression like (5.27), the nouns cannot occur in a direct construction with numerals. The numeral classifier *lêm* must be compulsorily used, presumably to make the noun *singulative*. On the contrary, in languages with a singular/plural system, such as English, the unmarked form is clearly singular in itself, so a numeral classifier serving to individuate those referents of nouns is unnecessary.

5.3 Other quantificational extra elements

In this section, the extra elements which are somewhat relevant to quantification are considered. Due to the lack of information however, the account of why these elements are used remains unsatisfactory. Note that although these elements may seem unrelated to counting, they have over time become obligatory constituents of the constructions. In any case, it might be more useful to keep them as quantificational extra elements rather than ignoring them. Further research on their historical origins is required for justification.

5.3.1 Relative clause marker

There exists a language in which a relative clause marker is involved in quantification. This is Kambera (Austronesian; Indonesia). In Kambera, if one counts two people or more, the numerals must be in a relative clause. For example,

- (5.28) *Kambera (Klamer 1998: 139)*
tau ma-dua
person RMS-two
'two persons/people' (lit. 'people that are two')

In this language, the subject relative clause marker (RMS) is used with the predicate. In counting *people*, the numeral *dua* 'two' is treated as the predicate of the relative clause, and the noun *tau* 'person' is treated as the head noun. It can be observed that the use of the relative clause marker in CNNCs is dependent on the numeral. The marker is used for the numeral *dua* 'two' onwards, but not for the numeral 'one'. As for the numeral 'one', the construction consisting of the prefix numeral *one* plus the noun is used instead, for example, *ha.atu tau* [one.CLF person] 'one person' (Klamer 1998: 136-141, 318-319). When non-human nouns are counted, the language uses numeral classifier constructions. In sum, the RMS was initially required for a syntactic reason, but now it is like a vestigial morpheme required in CNNCs with a restrictive use.

5.3.2 Numerical particle

In Maori (Austronesian, New Zealand) and some other Austronesian languages, the numerals must generally be preceded by the particle *e*. In Maori (Bauer 1993: 495 and passim) the function of the particle is unclear. According to Lynch (1998: 117-118), in many Oceanic languages (an Austronesian branch), numerals and quantifiers are stative verbs (i.e. verbs denoting a state of things, e.g. English *sleep*) as illustrated by Fijian, an Oceanic language spoken in Fiji island.

(5.29) *Fijian* (Lynch 1998:117)

a. *e moce na gone*
 3SG sleep DEF child
 'The child slept/is sleeping.'

b. *e dua na gone*
 3SG one DEF child
 '(There is) one child.'

It can be noticed that the numeral appears in the same slot with the stative verb and the particle *e* precedes the stative verb as well as the numeral. It is possible that the morpheme *e* has some function other than that of a personal pronoun.

Lynch points out that the numerals in many Oceanic languages show a defunct verbal prefix. The table in Lynch (1998: 118) is reproduced as shown in Table 5.1 below.

	Proto-Oceanic	Vinmavis	Lenakel
'two'	*rua	iru	kiu
'three'	*tolu	itl	kisil
'four'	*vati	ifah	kuvir
'five'	*lima	ilim	katilum

Table 5.1 Verbal prefixes for numerals in Oceanic languages (NB: the asterisks mark the reconstructed forms)

It can be observed that the roots of the numerals in Vinmavis are prefixed by *i* and in Lenakel by *k* (+vowel). The prefix *i* is a third person singular non-future verbal prefix and *k* (+vowel) is a third person non-singular verbal prefix. It is conjectured in Lynch (1998: 118) that these numerals were once stative verbs, and over time these prefixes have been attached to the numerals (cf. § 3.2.2.2 *lexicalization*), making the numerals become complex lexemes. Finally, the numerals have lost their verbal status.

Turning to Maori, the numerals are rather predicative (stative verb) than attributive (Bauer, Parker, and Evans 1993: 496). Also note that the stative verb in (5.30) is preceded by the particle *e*.

(5.30) *Maori* (Bauer, Parker and Evans 1993)

a. *e* *ono* *ngaa* *matapihi*
 NUMPCL *six* DEF.PL *window*
 ‘There are six windows.’ (lit. ‘The windows are six’.) (p.496)

b. *e* *moe* *ana* *te* *peepi*
 T/A *sleep* T/A DEF *baby*
 ‘The baby is sleeping.’ (p.421)

Considering the form and syntactic position of the particle *e* in (a) and (b), it may be conjectured that the particle *e* which is used for quantifiers in Maori is associated with the particle *e* as a tense/aspect marker. Therefore, the particle *e* might be required in CNNCs due to the verbal status of numerals. At present, the function of this particle in CNNCs seems to be unclear. The grammar states only that the particle is obligatory for numerals greater than one and used only with non-human nouns. The topic requires further research, though.

5.3.3 Accusative case marker

There are instances, though few, in which an accusative case is involved in CNNCs. An example is observed in Modern Standard Arabic. In this language, the quantified noun is marked for accusative case if modified by the numerals 11-99 even though the noun phrase does not occupy a position demanding an accusative case. This is relevant to quantification—that is, the choice of grammatical case depends on the numeral. However, the motivation for using the accusative case remains unclear. An example from Arabic is shown below.

- (5.31) *Arabic* (Holes 1995: 174)
- | | | |
|--|------------------|-----------------|
| <i>xams-a</i> | <i>ášarat-a</i> | <i>bint-an</i> |
| <i>five.F-ACC</i> | <i>ten.F-ACC</i> | <i>girl-ACC</i> |
| <i>'fifteen girls' ('fifteen as regards girl')</i> | | |

5.3.4 Oblique marker

The term *oblique* refers to any syntactic element which is neither a subject nor object of the verb (Matthews 1997: 253). In this thesis, the term *oblique* is used in CNNCs referring to a noun which has an attached preposition or case marker (i.e. a marker indicating the grammatical relation between the elements involved). The preposition or case marker denotes a certain relationship between the noun and the numeral. For example, in the noun phrase *two of the camels*, *of the camels* is oblique and *of* is an oblique marker. In the current data, there are various instances of oblique, namely *genitive singular marker*, *partitive singular marker*, *genitive plural marker*. These markers may also be represented by the preposition *of*.

The term *genitive* refers to the grammatical case which denotes that one noun is dependent on another noun. However, the noun which is marked for this case is basically a possessor of another noun. For example, in the Latin *cōnsulis equus* 'the consul's horse', *cōnsulis* is in the genitive form (and has the meaning *of consul*) denoting a possessor of a horse. The genitive forms may be different in number, for

example, Latin *manus* ‘of hand’ and *manuum* ‘of hands’. The former represents the *genitive singular* and the latter, the *genitive plural* (Blake 1994: 5-6).

As for the term *partitive*, generally the central function of the partitive case involves the idea of partiality. There are several related senses of partitive cases including the one which overlaps with the genitive case (see Koptjevskaja-Tamm 2006: 218 for detailed discussion about the definition). However, the sense which concerns CNNCs is that the partitive case is used to indicate that the noun which is marked for the partitive case denotes the whole set of another noun (or the equivalent). For example, in *three of the boys*, the phrase *of the boys* is partitive because it refers to the whole set from which a subset (three boys) is selected.

In the current data, several instances of CNNCs have the oblique as an extra element. Although some usages of the oblique in the data are not clearly understood, it seems likely that there is more than one possible account for the motivation for using the oblique in CNNCs. The first comes from Greenberg (1972: 181). As earlier mentioned (§5.2.4), Greenberg points out that a language in which a system of collectives exists may employ a grammatical device of individuation in a quantificational expression, since “a numeral cannot occur directly with a collective” (1972:179).

Among the alternatives is the use of one or more non-collectives in construction with the numeral and more loosely joined syntactically to the collective which is in apposition or is a dependent (partitive) genitive (1972: 181)

The reason why the partitive or genitive or the preposition *of* is used as a method of individuation is not focused on by Greenberg. However, it is likely that the oblique markers function as a part-whole (subset-set) linker. The *set* can be conceptualized as *collective* or *mass*. To count an individual member in the set, one needs to individualize the members in the set first (compare *two of them*, not **two them*). The counted members are then a part or a subset of the set. The representational structure in the speaker’s mind may be illustrated with the figure below.

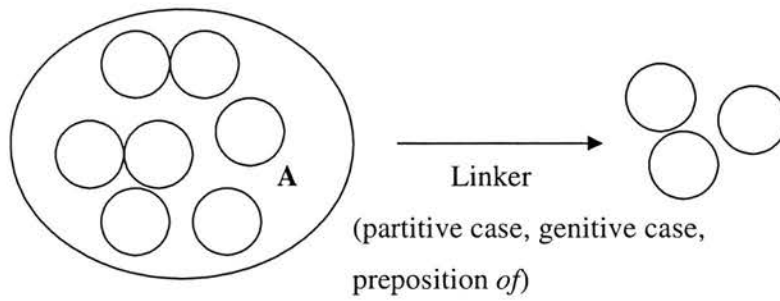


Fig. 5.2 Representational structure of counting collective references

In Fig. 5.2, the circle A represents a set of 7 members. When counting the members in the set, one needs to pull out the members from the set one by one (i.e. *individuation*). The individuation can be represented linguistically by the linker indicating the relationship between the set and the numerated members (i.e. the three small circles on the right). The linker as such can be the partitive case (the three circles are a subset of the circle A, or the genitive case (i.e. the three circles belong to the circle A). Both cases are semantically related and can alternatively be signified by the English preposition *of*.

The second account involves a *universal* in morphosyntactic properties of numerals (Corbett 1978). It is argued that in a language where numerals vary in syntactic category—that is, the low numbers behave adjectivally and the high numbers behave nominally—the insertion of the genitive or a preposition between numerical nouns and quantified nouns is required for the high numbers to avoid double nominatives (i.e. two nouns in adjacent position).

The final account is that the phenomenon is attributable to a historical accident, as happened in Old Russian where some nouns in the dual nominative and accusative forms looked identical in general with the *genitive singular* form, and hence the dual was reinterpreted as the *genitive singular*. Therefore the genitive singular form used in CNNCs is in fact (in this account) a remnant of dual number (Corbett 2000: 269).

5.3.5 Numeral marker for persons

There exists a language in which a marker referred to in the source as *numeral marker for persons* (PNUM) (Bauer, Parker, & Evans 1993: 496) is used in counting expression. The use of PNUM is evident in Maori (Austronesian; New Zealand). In this language, nouns denoting humans are prefixed with *toko-* if modified by the numerals 2-9. This is relevant to quantification—that is, the marker is used in counting expressions. However, the use of PNUM is regarded as traditional and is increasingly ignored. An example from Moari is shown below.

- (5.32) *Maori* (Bauer, Parker and Evans 1993: 496)
- | | | |
|-------------------------------------|-------------------|-----------------|
| <i>toko-rima</i> | <i>oona</i> | <i>tuaakana</i> |
| <i>PNUM-five</i> | <i>GEN.PL.3SG</i> | <i>brother</i> |
| <i>'He had five older siblings'</i> | | |

5.3.6 Double plural marker

Degema (Niger-Congo; Nigeria) is the only language in the current sample having a special marker referred to in the source as double plural marker (DPM) (Kari 2004: 209). In this language, most nouns are marked for number—singular and plural. It is noted in Kari (2004: 209) that “Non-partitive numerals are preceded by a double plural morpheme *mè* which is used in counting nouns designating certain plural entities.” An example is shown below.

- (5.33) *Degema* (Kari 2004: 209)
- | | | |
|-----------------------|------------|------------|
| <i>ímó</i> | <i>mé</i> | <i>va</i> |
| <i>child.PL</i> | <i>DPM</i> | <i>two</i> |
| <i>'two children'</i> | | |

5.4 Non-quantificational extra elements

There are some extra elements which are present in CNNCs but are not specifically required for quantification and hence do not play any role in classifying structural types of CNNCs. In particular there are various non-quantificational extra elements observed in the current data. This section illustrates only a few of them just to show that not every extra element is relevant to quantification or useful for the classification of structural types of CNNCs. These markers include gender markers, copulas, some case markers, and attributive markers, for instance. They are required by the general structural rules in the language.

5.4.1 Noun class (gender)

When numerals are in the position of a modifier, some languages show a gender marker in CNNCs. The gender is not relevant to quantification. The marker is required just because the language has a gender system, and gender is obligatorily present on the modifiers in addition to the numerals. For example, in Catalan (Indo-European; Spain), CNNC_{SG} requires a gender marker. Since gender marker is not regarded, CNNC_{SG} in Catalan is coded as {N,NUM}.

(5.34) *Catalan (Hualde 1992: 122)*

<i>una</i>	<i>noia</i>
<i>one.F</i>	<i>girl (F)</i>
<i>'one girl'</i>	

5.4.2 Copula

The copula refers to the grammatical element linking two elements, namely a subject and a predicative element, comparable to English *be*. Examples may be found in English: e.g. English *am* as in *I am cold* or *I am a doctor* where *am* is a linker between the subject (i.e. *I*) and predicative elements (i.e. *cold* and *a doctor*) (Matthews 1997: 77). Likewise, in Japanese, a copula is required to assist nouns and nominal adjectives (i.e. nouns used as adjectives) in forming a predicate (Iwasaki 2002: 42). For example,

(5.35) *Japanese (Iwasaki 2002: 38)*

nihonjin-no hito
Japanese-COP person
'a Japanese person'

In this case, *nihonjin* 'Japanese' is a noun and therefore requires the copula *no* to form a predicate. Note that a predicate here includes the attributive function. This is perhaps because the language does not allow double nominatives, so the copula is a device to change the noun into an adjective.

All Japanese numerals require the copula when used as a modifier for the quantified noun. This situation is different from the use of the oblique to avoid double nominatives (cf. §5.3.4) in which quantification plays a role. That is to say, the oblique is required for high numerals, but not in low numerals. On the contrary, in Japanese the copula *no* is not required due to quantification at all, but rather for a syntactic reason.

5.4.3 *Attributive*

In some languages in the current data, the numerals as a modifier require an attributive morpheme to form them as an attributive. Examples are Kolyma Yukaghir (Yukaghir; Siberia, Russia), and Tok Pisin (Creoles and Pidgins; Papua New Guinea).

(5.36) *Kolyma Yukaghir (Maslova 2003a: 82)*

irk-in aŋd'e
one-ATTR eye
'one eye'

In Yukaghir, cardinal numerals have attributive and predicative forms (Maslova 2003a: 82). So, the attributive marker is used to show the attributive function of the numeral in a sentence. The attributive markers are not specifically employed for numerals. Rather any linguistic form requiring attributive function may require an attributive form. Overall, there is no quantificational motivation for the use of the attributive marker.

(5.37) *Tok Pisin* (Verhaar 1995: 29)

<i>wan-pela</i>	<i>boi</i>
<i>one-ATTR</i>	<i>boy</i>
<i>'one boy'</i>	

In Tok Pisin, it is noted that the suffix *-pela* is used to form a limited number of modifiers to nouns, some pronouns, and a variety of numerals (Verhaar 1995: 12). Since it is not used only with numerals, the quantificational motivation is not clear. Both languages are therefore categorized as {N,NUM}.

5.4.4 *Switch reference*

The switch reference marker is used to disambiguate subjects of the two verbs in complex sentences consisting of more than one clause. Switch reference marking can be observed in Koasati (Muskogean; US, Alabama) and other Muskogean languages. In this language, numerals are verbs and the head noun of the numeral is therefore a subject of the verb. If the clause consists of a numeral and noun embedded in the complex sentence, the switch reference marker may be required. Without the reference marker, it may cause confusion—that is, confusion over whether the two verbs belong to the same subject or different subjects. For example,

(5.38) *Koasati* (Kimball 1991: 358)

<i>ná:ni-ha</i>	<i>pokkó:l</i>	<i>awáh</i>	<i>tóklo-n</i>	<i>hí:ca-li-:s</i>
<i>man-PL</i>	<i>ten</i>	<i>and</i>	<i>two-SW</i>	<i>see-ISS.PST</i>
<i>'I just saw twelve men.'</i>				

The affix *-n* is a reference marker, indicating that the subject of the numeral verb 'twelve' is not the same as the subject of the verb 'see'. Without it, it might have other interpretations, such as 'the twelve men saw me' for instance. In any case, the switch reference is not regarded as a quantificational extra element. This is because this switch reference marker is required because of sentence structure. If there is no other subject, the switch reference marker might not be needed, and most importantly there is no quantificational motivation involved.

5.5. Conclusion

This chapter has set out a criterion that the project has used for selecting extra elements in establishing types of CNNCs. The criterion is that the extra elements must show a certain quantificational motivation. As a result, we have two major extra elements—namely, number markers and numeral classifiers. There are a few other quantificational extra elements used as well, such as relative clause markers, numerical particles, accusative case markers and oblique markers. There are many non-quantificational extra-elements observed in CNNCs, for example, noun class markers, copulas, attributives, and switch reference markers. These are not taken into consideration when establishing types of CNNCs. The said quantificational motivation of the number markers and numeral classifiers stems from the vagueness of nouns with reference to inherent number. These nouns are called *general nouns*. The emergence of the quantificational extra elements may be attributable to the principle of distinctness in language. Another quantificational motivation for the emergence of extra elements is concerned with the grammatical status of numerals. In some languages the high numerals behave nominally and so require syntactic elements to form CNNCs. Overall, these extra elements show that quantification brings about the grammatical complexity in language.

6 CNNCs Across Languages

This chapter illustrates various *language types* as they are classified according to CNNCs along with their frequencies and geographical distribution. A task such as this is referred to in Croft (2003:1) as *typological classification*. The notion *language type* refers to types of languages which are classified on the basis of their *structural types* under consideration (i.e. in the present context, types of structural patterns seen in CNNCs). For example, based on the order of cardinal numerals with respect to the noun they modify, English is an example of a language in which the numeral precedes the noun, e.g. *two boys*. Very often it is difficult to identify a language as a particular type because languages may employ more than one structural type in a given construction. For example, certain English constructions can have the adjective-noun order (i.e. the adjective precedes the noun) as in *military court* or the noun-adjective order (i.e. the noun precedes the adjective) as in *court martial*. However, the most practical solution is to look only at the *basic type*—that is, the most frequent structural type used in general contexts, not those restricted to a specific context nor used only with a special meaning (Croft 2003: 42-43).

Like most linguistic constructions, there often exist two or more structural types of CNNCs in a single language. In typologizing languages with respect to the means by which they form CNNCs, only the basic structural patterns are considered. Nevertheless, some sources do not provide evidence of the basic structural patterns. In this case, the patterns taken from the relevant examples available in the sources were assumed to represent the basic types for the reason that the examples were randomly chosen. In some languages, a mixture of structural types is allowed with none being primary, for example, when the nouns in the language are generally *optionally* marked for plural. Those languages are classified as *Mixed*. The mixed type also includes cases where the structural patterns observed used complementarily, for example, one structural pattern is used for human nouns and another for non-human nouns. In such cases, it is difficult to

assign the language to a particular type. Note that in some languages of mixed type, it is probable that one of the structural patterns observed is dominant over the others in actual use, but due to the absence of evidence, all structural types observed were assumed to be used equally.

In addition to the basic structural type, most languages have other structural types used as *subsidiary* strategies for forming CNNCs. The subsidiary types are used with constraints—that is, their occurrences are limited to a small set of nouns or used in a special context. Refer to the introductory section of Appendix 3 for the examples of restrictive cases. Because only the basic types are considered in typologizing languages, these subsidiary types are not taken into consideration here. However, the subsidiary types are reported in the inclusive tables of $CNNC_{SG}$ and $CNNC_{NSG}$ offered in Appendix 1. They are represented with the symbol @.

The organization in this chapter is inspired by the well-organized illustration of typological data in the World Atlas of Language Structures (WALS) (Haspelmath, Dryer, Gil, Comrie 2005). This chapter is divided into two major parts. The first part deals with a typological description of $CNNC_{SG}$, while the second part deals with a typological description of $CNNC_{NSG}$. Each part begins with establishing *language types* of CNNCs, followed by a description of each language type. Each language type is illustrated by examples taken from different languages in the current sample. The typological description is accompanied by maps (a single world map and small blow-up maps of certain zoomed areas where the dots are much too dense). The maps show the geographical distribution of those language types corresponding to different colored dots. Since the subsidiary types are not taken into consideration in typologizing languages, they are therefore not shown on the maps. Both world maps and blow-up maps of certain zoomed areas are generated with the *Interactive Reference Tool of the World Atlas of Language Structures (WALS)*, developed by Hans-Jörg Bibiko (2005).

6.1 Types of CNNC_{SG}

This section describes the language types of CNNC_{SG}. This construction was examined on the basis of 175 languages. On the basis of the structural patterns of CNNC_{SG} used as a basic type in a given language, 6 values are established, as shown in Table 6.1 below. For a list of languages in which these types are evident, the reader is referred to Appendix 2.2. Also, note that the numbered types in Table 6.1 do not correspond to those in Table A in Appendix 1. This is because the types shown in Table 6.1 are *language types*, while those shown in Table A in Appendix 1 are *structural types*.

Type	Number of languages
1. {N,NUM}	117
2. {N,NUM,SG }	10
3. {N,NUM,CLF}	39
4. {N,NUM} + {N,NUM,CLF}	5
5. Other	2
6. None	2
Total	175

Table 6.1 Language types with regard to CNNC_{SG} and their frequencies
NB: (+) = a mixture of constructions (presumably, with none primary)

The first type ({N,NUM}) includes languages in which the CNNC_{SG} simply consists of a noun and the numeral *one* without any additional morpheme, as in English *one man*. Note again that the order of the elements in the set notations is disregarded.

The second type ({N,NUM,SG}) includes languages which form their CNNC_{SG} by attaching the singular marker to the noun and/or the numeral *one*, as illustrated by the example in (6.1) from Alutor (Chukotko-Kamchatkan; Russia). In this language the morpheme *-ŋa* is suffixed to the noun, indicating the singular number (and simultaneously case). Some languages may show agreement for singular number (and simultaneously class) both on the noun and the numeral, as shown in Lunda (Niger-

Congo; Democratic Republic of Congo) in (6.2). The double singular marker as such is also represented by the notation {N,NUM,SG}, although not {N,NUM,SG,SG}.

(6.1) *Alutor* (Kibrik, Kodzasov and Muravyova 2004: 371, 508)

<i>ənnan</i>	<i>rara-ŋa</i>
<i>one</i>	<i>house-NOM.SG</i>
<i>'one house'</i>	

(6.2) *Lunda* (Kawasha 2003: 124)

<i>mu-ntu</i>	<i>wu-mú</i>
<i>I-person</i>	<i>I-one</i>
<i>'one person' (Class mu-/wu- = singular number)</i>	

The third type ({N,NUM,CLF}) includes languages where the $CNNC_{SG}$ is expressed by the combination of the numeral *one*, a noun, and a numeral classifier, illustrated by the example in (6.3) from Mandarin (Sino-Tibetan; China). Also, a couple of languages of this type show double numeral classifiers, such as Piaroa (Sàliban; Venezuela), illustrated in (6.4). The double numeral classifiers as such are also represented by {N,NUM,CLF}, although not {N,NUM,CLF,CLF}.

(6.3) *Mandarin* (Lin 2001: 107)

<i>yí</i>	<i>gè</i>	<i>jīdàn</i>
<i>one</i>	<i>CLF</i>	<i>egg</i>
<i>'one egg'</i>		

(6.4) *Piaroa* (Krute 2003:144)

<i>kurod-æ</i>	<i>hid-æ-tetæ</i>
<i>machete-CLF</i>	<i>one-CLF-one</i>
<i>'one machete'</i>	

The fourth type includes languages which exhibit a mixture of {N,NUM} and {N,NUM,CLF}, as may be seen from the example in (6.5a-b) from Nicobarese (Car) (Austro-Asiatic; Nicobar Islands, India). This pattern of mixture is relatively frequent compared to other patterns of mixture and therefore is considered separately.

The example in (6.5a) illustrates {N,NUM}, while the example in (6.5b) illustrates {N,NUM,CLF}.

(6.5) *Nicobarese (Car) (Braine 1970: 113)*

a. {N,NUM}

héŋ kúk
one coconut
 'one coconut'

b. {N,NUM,CLF}

héŋ nɔŋ ʔá:p
one CLF canoe
 'one canoe'

The fifth type includes the languages of other possibilities and hence is classified as *OTHER*. One is the combination of a noun and a singular marker or a singulative marker (cf. §5.1.2) without the overt numeral *one*. Mohawk (Iroquoian; Canada, United States) is the only language in the current sample employing this structural pattern as a primary strategy of forming $CNNC_{SG}$ as in (6.6a). The language employs the structural pattern of {N,NUM} as in (6.6b) as well, though the pattern is regarded as marked (Marianne Mithun, personal communication).

(6.6) *Mohawk (Bonvillain 1973: 235)*

a. {N,SG}

ska-háŋselah
SG-lamp/light
 'one lamp, light'

b. {N,NUM}

v́hska oháhaselaʔ
one lamp/light
 'one lamp, light'

The pattern of {N,SG} may be regarded as a marginal case of CNNC_{SG}. Although structurally it does not contain the numeral *one*, semantically the construction expresses the numerosity *one* of the referent. Also, considering the forms of the singular or singulative markers, it seems likely that the singular or singulative markers in many languages are etymologically related to the numeral *one*. This probably suggests a grammatical change from the numeral *one* to the singular or singulative markers. Moreover, {N,SG} may also exhibit an intermediate phase in the diachronic route between {N,NUM} and {N,NUM,SG}, as evidenced in Pame (Oto-Manguean; Mexico) (cf. §8.2.1.2). Thus, it is rather useful to include {N,SG} in the classification of structural types of CNNC_{SG}.

The other possibility is the co-existence of {N,NUM,SG} and {N,NUM,OBL,SG} in a language. This mixture is observed in Berber (Ayt Seghrouchen Middle Atlas) (Afro-Asiatic; Morocco). The example in (6.7a) illustrates {N,NUM,SG}, while the example in (6.7b) illustrates {N,NUM,OBL,SG}.

(6.7) *Berber (Ayt Seghrouchen Middle Atlas) (Penchoen 1973: 25)*

a. {N,NUM,SG}

<i>yun</i>	<i>u-ryaz</i>
<i>one</i>	<i>SG-man</i>
<i>'one man'</i>	

b. {N,NUM,OBL,SG}

<i>yut</i>	<i>n-t-məttut</i>
<i>one</i>	<i>of-SG-woman</i>
<i>'one woman'</i>	

In fact, the two structural patterns are used rather complementarily. The pattern of {N,NUM,SG} is used with nouns beginning with a vowel (i.e. masculine nouns). On the other hand, the pattern of {N,NUM,OBL,SG} is used with nouns beginning with a consonant (i.e. feminine nouns) as well as those unberberized nouns borrowed from Arabic (Penchoen 1973: 25).

Finally, just two of the languages in the survey lack the numeral *one* and subsequently do not have CNNC_{SG}. These languages are Pirahã (Mura; Brazil) and Wari' (Chapacura-Wahan; Brazil). Although they do not have the numeral *one* proper, Pirahã uses the word *hói* 'small size' or 'small amount', and Wari' uses the word *xica* 'alone' to convey the concept of the number *one*, as shown below.

(6.8) *Pirahã* (Everett 2005: 623)

<i>tiobáhai</i>	<i>hói</i>	<i>hii</i>	
<i>child</i>	<i>small</i>	<i>PRED</i>	

'small child/child is small/one child'

(6.9) *Wari'* (Everett and Kern 1997: 348)

<i>xica'</i>	<i>pe</i>	<i>na</i>	<i>tarama'</i>
<i>alone</i>	<i>be.at.SBJ</i>	<i>3SG.RP/P</i>	<i>man</i>

'There is one man.' (lit. 'The man is alone.')

The structural types just described so far all are employed as the basic means of forming CNNC_{SG} in the world's languages. They are also, however, employed as a subsidiary mode in some languages. For example, {N,NUM,CLF} is a subsidiary mode in Turkish (Altaic; Turkish), a language in which {N,NUM} is the basic type of CNNC_{SG} (Kornfilt 1997: 271).

In addition to the types illustrated above, there are a few other structural types of CNNC_{SG} which are employed only as a subsidiary mode due to their restricted use. Each is extremely rare, being represented by only a couple of instances. The first of these is the use of the word meaning 'unit' instead of the numeral *one* in Somali (Afro-Asiatic; Somalia). An example is given in (6.10) below.

(6.10) *Somali* (Saeed 1999: 58)

<i>xabbàd</i>	<i>līn</i>	<i>áh</i>
<i>unit</i>	<i>orange</i>	<i>be</i>

'one orange' (lit. 'a unit that is orange')

The use of the word meaning ‘unit’ may be comparable to the use of numeral classifiers without the numeral *one* or {N,CLF}(see the next type below). That is to say, both are used to form or unitize the quantified noun (which is conceptually vague in number) into one unit (cf. Greenberg 1972: 183) and the numeral *one* is no longer needed for the reason that the concept of ‘unit’ also expresses the quantity *one* implicitly.¹⁴ This structure is assumed to be non-dominant in the language because it is not mentioned at all in other Somali grammars (for example, Orwin 1995: 68). This type is regarded as a *marginal* case of CNNC_{SG}, since the word meaning ‘unit’ is not the numeral *one*, and hence one of the core constituents is missing. As suggested by Bernard Comrie (personal communication), we may check whether the element in question is a numeral by simply asking the native speakers with the ‘how many?’ question. In this case, if we ask native speakers how many oranges there are, they might not say *xabbàd* ‘unit’ even though there is only one orange.

The second subsidiary type is a combination of a noun and a numeral classifier with no morpheme at all expressing the numerosity *one* (or {N,CLF}). Vietnamese (Austro-Asiatic; Vietnam) and Thai (Tai-Kadai; Thailand) provide examples of this structure. Like the case of Somali, this type is also regarded as a marginal case of CNNC_{SG}, since the construction does not have the core element, namely the numeral *one*. Although the numeral *one* is not present in the constructions, these constructions express the individuality of the noun. They therefore express the numerosity *one*, as shown in (6.11) and (6.12).

(6.11) *Vietnamese (Bisang 1996: 541 in Bisang 1999: 146)*

<i>tôi</i>	<i>mua</i>	<i>quả</i>	<i>cam</i>
<i>1SG</i>	<i>buy</i>	<i>CLF</i>	<i>orange</i>

‘*I buy the orange. / I buy an orange.*’

¹⁴ The situation can be compared to the use of the quantifier in colloquial Thai (own knowledge). When one says *ná:m kə̀w* [water glass], this expression means ‘one glass of water’. The noun ‘glass’ functions as a unitizer, comparable to the function of the noun ‘unit’ in Somali. Therefore, the noun ‘unit’ seems to be used to express the individuality of the noun.

(6.12) *Thai (own knowledge)*

tó *tua*
table *CLF (lit. 'body')*
'one table'

The third subsidiary type is a combination of the numeral 'one' and a numeral classifier (or {NUM,CLF}). This type is observed in many numeral classifier languages including Nepali (Indo-European; Nepal, see (6.13)).

(6.13) *Nepali (Hutt and Subedi 2003: 56)*

ek *chin*
one *moment*
'one moment'

Generally, in numeral classifier languages {NUM,CLF} is used only when the noun and the numeral classifier share the same form. In fact, the noun itself is used as a numeral classifier, and then the noun need not be repeated, since the meaning of the noun is clear from the numeral classifier. The noun should be analysed as a numeral classifier rather than a head noun when it is positioned in the same syntactic slot with numeral classifiers in the language (Bernard Comrie, personal communication). These nouns are mostly nouns meaning 'human' or denoting units such as 'day' and 'month'.

The fourth type is composed of a noun, a numeral and a non-plural marker (or {N,NUM,NPL}), as illustrated in Imonda (Border; Papua New Guinea). The non-plural marker has already been mentioned in §5.1.4.3, so it will not be discussed further here. The example is illustrated here again for convenience.

(6.14) *Imonda (Seiler 1985: 62)*

toad-ianèi *mugasl*
boy.PL-NPL *one*
'one boy'

The final subsidiary type is a combination of a noun, the numeral *one*, a singular marker, and a numeral classifier (or {N,NUM,CLF,SG}) as in Ejagham (Niger-Congo; Cameroon, Nigeria) the only language in the sample where this type is observed. The singular marker is a portmanteau morpheme encoding class and a singular number simultaneously. For further discussion on this structural type, the reader is referred to §8.4.

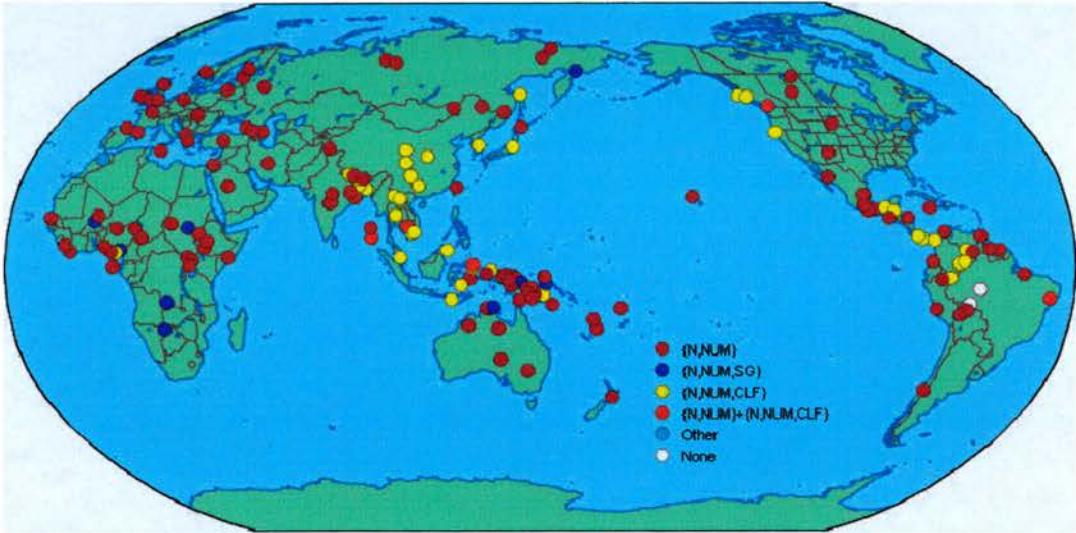
(6.15) *Ejagham* (Watters 1981: 469)

<i>ê-rǎm</i>	'	<i>í-čókùd</i>	<i>jǎ-d</i>
<i>CL-CLF</i>	<i>GEN</i> ¹⁵	<i>CL-orange</i>	<i>CL-one</i>
'one orange' (Class <i>ê-</i> = singular number)			

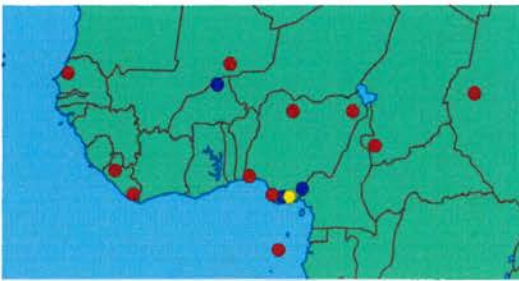
¹⁵ Aikhenvald (2000: 99) refers to the tone as 'genitive linker'.

6.2 Geographical distribution of $CNNC_{SG}$

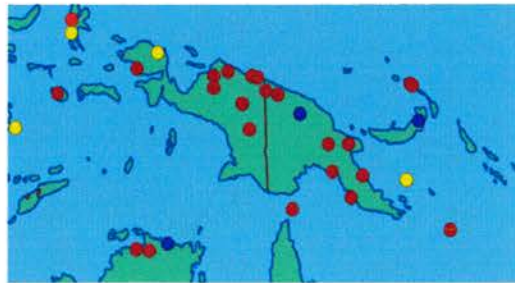
In this section the focus is on the worldwide patterns of occurrence of language types with respect to $CNNC_{SG}$, as depicted in Map 6.1 below.



Map 6.1 Geographical distribution of $CNNC_{SG}$



Map 6.2 Western Africa (zoomed)



Map 6.3 New Guinea and Northern Australia (zoomed)

Map 6.1 basically divides the languages of the world into two major types with regard to $CNNC_{SG}$, namely $\{N,NUM\}$ (shown in red) and $\{N,NUM,CLF\}$ (shown in yellow), with very few other possibilities. As is evident from the map, $\{N,NUM\}$

spreads through almost all regions, greatly outnumbering all other types combined. This suggests unsurprisingly that to form $CNNC_{SG}$, the world's languages mostly prefer the simple structural pattern which requires no additional element other than the necessary elements expressing the numerosity *one* and the noun referent. The largest areas in which {N,NUM} languages are absent are East and Southeast Asia, the areas where the main concentration of numeral classifier languages is to be found, for example, the Sino-Tibetan and Tai-Kadai languages. The languages with {N,NUM,CLF} are fairly common around the Pacific Rim (i.e. East Asia, Southeast Asia including New Guinea and the surrounding areas, and the west coasts of North America and South America) and South Asia with various degrees of frequency. The geographical distribution of {N,NUM,CLF} in relation to degrees of frequency is further discussed in §6.4.

The next type is {N,NUM,SG} (shown in blue). The singular marker here refers to a morphological device for expressing singularity. The marker can be a separate single morpheme or a portmanteau morpheme that fuses a singular marker with other grammatical categories such as noun class or case. Languages where this type is the basic strategy of forming $CNNC_{SG}$ are relatively rare—only 10 such languages are observed in the sample (see Appendix 2.2 for a list of languages of this type). Six of these are present in Africa (e.g. Lunda, spoken in Democratic Republic of Congo). These languages are genetically related—that is, they are all affiliated to the Niger-Congo family (albeit different genera) which is notable for having extensive noun class systems. The noun class markers in these languages simultaneously encode grammatical numbers (singular and plural) (Williamson and Blench 2000: 12). In addition, a couple of instances are also found in the Australia-New Guinea region. Like the six languages in the Niger-Congo family, the singular marker in Yimas (Lower Sepik-Ramu; Papua New Guinea) and Ndjébbana (Australian; Northern Territory) is expressed by a portmanteau morpheme encoding both noun class and singular number. Another two languages are found to have {N,NUM,SG}, namely Sulka (isolate; Papua New Guinea) and Alutor (Chukotko-Kamchatkan; Siberia). The noun class system is absent in Sulka (Nichols 1992: 298-299) and there is no evidence indicating whether Alutor possesses noun class systems or not. Still, the singular marker in Alutor is expressed together with

case. Apart from this, this structural pattern is observed as a subsidiary mode in a few instances from Persian (Indo-European; Iran), Burushaski (isolate; Pakistan), and Pame (Oto-Manguean; Mexico).

Next, the languages which are classified as a mixture of {N,NUM} and {N,NUM,CLF} (shown in orange) are not unusual among the numeral classifier languages. Unsurprisingly, the mixture of the two constructions is often found at the peripheral areas where numeral classifier languages are found, typically the southernmost of Southeast Asia. Examples are Tidore (West Papuan; Indonesia) and Nicobarese (Car) (Austro-Asiatic; India (Nicobar Islands)). This type of mixture can also be found in the areas where languages using both constructions are adjacent, for example, Halkomelem (Salishan; Canada). Refer also to Map 6.4 for geographical distribution of this type. Although a language classified as mixed is a problem for typologization, such mixtures reflect the transitional stage of types in the language (Croft 2003: 44). For example, a mixture of {N,NUM} and {N,NUM,CLF} may exhibit a transitional stage between a language type of {N,NUM,CLF} and {N,NUM} or the other way round. The phenomenon of type change may not be particularly interesting, as it is common in languages and can occur in other parts of grammatical systems. Nevertheless, as Croft (2003: 245) points out, the interesting question is how we can identify the intermediate stages during the change from one system to another system. In the case of CNNCs, for example, the intermediate stage between {N,NUM} and {N,NUM,CLF} may involve a degree of obligatoriness in the numeral classifier in CNNCs and constraints on their occurrence. According to Aikhenvald (2000: 121), in a language with “incipient systems of numeral classifiers” (i.e. the stage where it is assumed that the numeral classifier system begins to develop), such as Omani, an Afro-Asiatic spoken in Oman, the classifiers are used optionally and are limited to some classes of nouns. The diachronic routes of types of CNNC_{SG} are discussed along with the causal factors driving the changes in Chapter 8.

There are two languages on Map 6.1 classified as *OTHER*. The first is {N,SG} which is observed in Mohawk (Iroquoian; Canada, United States) in North America. As mentioned earlier, Mohawk is the only language in the data using {N,SG} as a primary

mode of $CNNC_{SG}$. In fact, $\{N,SG\}$ is not rare in the world's languages, it is also observed as a non-primary mode in another three languages from Eurasia and North America, namely Burushaski (isolate; Pakistan), Persian (Indo-European; Iran), and Pame (Pamean; Mexico) (cf. Appendix 1, Table A).

In addition to the mixture of $\{N,NUM\}$ and $\{N,NUM,CLF\}$, there exists a language showing a mixture of $\{N,NUM\}$ and $\{N,NUM,OBL,SG\}$ which is grouped in the category *Other*, since there is only one language observed. This mixture is evident in Berber (Ayt Seghrouchen Middle Atlas) (Afro-Asiatic; Morocco).

Finally, as already mentioned in §4.4, there are only two languages, namely Wari' and Pirahã in the current sample, which lack the numeral *one* and lack $CNNC_{SG}$ accordingly. Both are spoken in Brazil, South America, the continental area which is remarkable for having only low-valued numerals.

6.3 Types of $CNNC_{NSG}$

On the basis of 233 languages, it is possible to establish 11 language types with respect to $CNNC_{NSG}$, as shown in Table 6.2 below. For a list of languages where these types are evident, the reader is referred to Appendix 2.3. Also, note again that the numbered types in Table 6.2 do not correspond to those in Table B in Appendix 1. This is because the types shown in Table 6.2 are *language types*, while those shown in Table B in Appendix 1 are *structural types*.

Type	Number of languages
1. {N,NUM}	78
2. {N,NUM,NSG }	36
3. {N,NUM,CLF}	33
4. Other	10
5. {N,NUM}+{N,NUM,NSG}	33
6. {N,NUM}+{N,NUM,CLF}	3
7. {N,NUM,CLF}+{N,NUM,CLF,NSG}	8
8. {N,NUM,(NSG)}+{N,NUM,OBL,SG/NSG}	5
9. {N,NUM,(NSG)}+{N,NSG}	12
10. Mixed	13
11. None	2
Total	233

Table 6.2 Language types with regard to $CNNC_{NSG}$ and their frequencies

The first type includes the languages where the $CNNC_{NSG}$ consists of a noun and a numeral greater than one (or {N,NUM}) (hereafter in this section, the numerals which are mentioned refer to numerals greater than one). Kpelle (Niger-Congo; Liberia) is an example of this type.

(6.16) *Kpelle* (Welmers 1973: 294)

<i>pére</i>	<i>feere</i>
<i>house</i>	<i>two</i>
<i>'two houses'</i>	

The second type comprises the languages in which the $CNNC_{NSG}$ is expressed by means of a noun combined with a numeral plus an obligatory non-singular marker (or {N,NUM,NSG}), for example, English *two dog-s*. Some languages may have double non-singular markers. Once again, Lunda (Niger-Congo; Democratic Republic of Congo) provides a clear example, as shown below.

(6.17) *Lunda* (Kawasha 2003: 124)

<i>a-ntu</i>	<i>a-yedí</i>
<i>II-person</i>	<i>II-two</i>
<i>'two people'</i> (Class <i>a-</i> = plural number)	

The third type includes languages where the $CNNC_{NSG}$ is accompanied by a numeral classifier in addition to the core constituents (or {N,NUM,CLF}). An example is given in (6.18) from Bai (Sino-Tibetan; China)

(6.18) *Bai* (Wiersma 2003: 671)

<i>sú^f</i>	<i>ŋv</i>	<i>ts^huɛ²</i>
<i>book</i>	<i>five</i>	<i>CLF</i>
<i>'five books'</i>		

The fourth are languages showing some other rare possibilities of $CNNC_{NSG}$, hence referred to as *Other*. All types are represented by only a couple of languages. There are six possibilities. The first is the combination of a numeral, a noun and a singular marker plus an oblique case marker (or {N,NUM,OBL,SG}). The oblique case marker sometimes also expresses the singular number as a portmanteau morpheme. There are two languages from the same genus possessing this type, namely Finnish (Uralic; Finland) and Estonian (Uralic; Estonia). The following example comes from Finnish.

- (6.19) *Finnish (Nelson 1998:46)*
yhdeksän omena-a
nine apple-PRTV.SG
'nine apples'

The second possibility is found in the languages where the $CNNC_{NSG}$ is expressed by means of a noun and a non-singular marker without an overt free numeral (or {N,NSG}). Three languages show this type, namely Bunuba (Australian; Western Australia), Diyari (Australian; South Australia) and Waorani (Waorani; Ecuador). An example is taken from Diyari in which the dual marker can be used instead of the numeral word 'two', as in (6.20). The non-singular marking morphemes are used as low-valued numerals, typically *two* and *three*. This type also includes the use of the third person dual pronoun, as in (6.21) from Waorani.

- (6.20) *Diyari (Austin 1981: 128)*
mankada-wuḷa
girl-DU
'two girls',

- (6.21) *Waorani (Peeke 1994: 269)*
tōdīya-da
sibling-3DU
'two brothers'

The third possibility is the combination of a noun, a numeral and the numerical particle (NUMPCL) (cf. § 5.3.2). (or {N,NUM,NUMPCL}). This type is found in Maori (Austronesian; New Zealand, see (6.22)), and Tuvaluan (Austronesian; Oceanic), as a subsidiary type.

- (6.22) *Maori (Bauer 1993: 262)*
eenei waka e rua
these canoe NUMPCL two
'these two canoes'

The fourth possibility is the combination of a noun, a numeral and a numeral classifier and a non-singular marker (or {N,NUM,CLF,NSG}). This type is found in Tariana (Arawakan; Brazil) and Tsimshian (Coast) (Penutian; Canada, United States). An example is taken from Tariana.

- (6.23) *Tariana* (Aikhenvald 2002:99)
- | | |
|--------------------------|------------------|
| <i>kehpuni-dapana-pe</i> | <i>panisi-pe</i> |
| <i>four-CLF-PL</i> | <i>house-PL</i> |
| <i>'four houses'</i> | |

The fifth possibility is the combination of a noun, a numeral and an inverse number marker (§5.1.4.4). This type is found in Jemez (Kiowa-Tanoan; New Mexico), as in (6.24).

- (6.24) *Jemez* (Yumitani 1998: 150)
- | | | |
|-------------------------------|----------------|--------------------------------|
| <i>wî</i> | <i>m̥sá-š</i> | <i>ṭil-kʷâ</i> |
| <i>two</i> | <i>cat-INV</i> | <i>TR.1SG.3DU-put.down/PFV</i> |
| <i>'I put down two cats.'</i> | | |

The sixth possibility is the combination of a noun, a numeral and a singular marker. This type is found in Zuni (isolate; New Mexico), as in (6.25).

- (6.25) *Zuni* (Nichols 1997: 12)
- | | | | |
|------------------------------|--------------|------------------|--------------|
| <i>ho'</i> | <i>ha'i</i> | <i>'e'ni-nne</i> | <i>'illi</i> |
| <i>1SG.NOM</i> | <i>three</i> | <i>belt-SG</i> | <i>have</i> |
| <i>'I have three belts.'</i> | | | |

The fifth type is the mixed type of {N,NUM} and {N,NUM,NSG} (or {N,NUM}+{N,NUM,NSG}). For example, in Hausa (Afro-Asiatic; Niger, Nigeria), animate nouns are marked for plural, while inanimate nouns are optionally marked in this respect (Schuh 1991: 1). This language therefore has two possible structural patterns, as shown in (6.26).

(6.26) *Hausa* (Schuh 1999: 1)

a. {N,NUM,NSG}

shanu *biyu*

cow.PL *two*

'two cows'

b. {N,NUM}

gida *goma*

house *ten*

'ten houses'

The sixth type is the mixed type of {N,NUM} and {N,NUM,CLF} (or {N,NUM}+{N,NUM,CLF}). Khmer (Austro-Asiatic; Cambodia) is an example of such a language, as in (6.27). The pattern of {N,NUM,CLF} is used in careful speech or written language, while the pattern of {N,NUM} is used in spoken language (Jacob 1965: 145).

(6.27) *Khmer* (Jacob 1990: 84)

a. {N,NUM,CLF}

mənùs *bu:ən* *nèək*

mankind *four* *CLF*

'four people'

b. {N,NUM}

chkae *bɔy*

dog *three*

'three dogs'

The seventh type is represented by the languages where the $CNNC_{NSG}$ is the mixture of {N,NUM,CLF} and {N,NUM,CLF,NSG}. Barasano (Tucanoan; Colombia) is an example.

(6.28) *Barasano (Jones and Jones 1991: 59)*

a. {N,NUM,CLF,NSG}

hɯa-rahe *kōbe-rahe-ri*

two-CLF *metal-CLF-PL*

'two metal cans' (NB: *rahe* 'cylinder', used as CLF)

b. {N,NUM,CLF}

gɯ̄a-hāl *hɯa-hāl*

stone-CLF *two-CLF*

'two stone slabs'

The eighth type includes the languages where the $CNNC_{NSG}$ is the mixture of {N,NUM} and/or {N,NUM,NSG} and a combination of a noun, a numeral, an oblique case marker plus a singular marker or a non-singular marker (or {N,NUM,OBL,SG/NSG}). The oblique case marker sometimes also expresses the (non-) singular number as a portmanteau morpheme. Examples are taken from Russian (Indo-European; Russia) and Welsh (Indo-European; Wales). In Russian, the structural pattern of {N,NUM} is used with the numerals ending with 1 (e.g. 21), while the structural patterns of {N,NUM,OBL,SG} and {N,NUM,OBL,NSG} are used with the numerals 2-4, and the numerals greater than 4 respectively (Neidle 1988: 90-95), as shown in (6.29). As for Welsh, the structural pattern of {N,NUM} is generally used with the low-valued numerals, while {N,NUM,OBL,NSG} is used with high-valued numerals (Thorne 1993: 149), as in (6.30).

(6.29) *Russian (Neidle 1988)*

a. {N,NUM}

dvadcat' *odin* *student*

twenty- *one* *student*

'twenty-one students' (p. 102)

b. {N,NUM,OBL,SG}
dva dnja
two day.GEN.SG
 'two days' (p. 95)

c. {N,NUM,OBL,NSG}
pjat' mal'čikov
five boy.GEN.PL
 'five boys' (p. 90)

(6.30) *Welsh* (Thorne 1993: 149)

a. {N,NUM}
dwy ferch
dwy girl
 'two girls' (p. 31)

b. {N,NUM,OBL,NSG}
naw o ddynion
nine of man.PL
 'nine men' (p. 149)

The ninth type is the mixture of {N,NUM} and/or {N,NUM,NSG} and {N,NSG}. An example of {N,NUM}+{N,NUM,NSG}+{N,NSG} is taken from Wambaya (Australian; Northern Territory, see (6.31)).

(6.31) *Wambaya* (Nordlinger 1998)

a. {N,NUM}
murrgunji alaji
three.CL boy.CL
 'three boys.' (p. 78)

b. {N,NSG}
darrangu-wulu
stick-DU
 'two sticks.' (p. 74)

- c. {N,NUM,NSG}
 gujarrawulu alag-ulu
 two child-DU
 'two kids.' (p. 75)

The tenth type is a group of languages classified as mixed. There are several possibilities with no particularly systematic patterns. Some languages show a mixture of more than 2 types of CNNC_{NSG}. As noted earlier, in each possibility, some types might be more dominant than others, but due to the lack of evidence, the types are treated as equal. However, what is interesting about the mixed group is that these languages show a variety of structural patterns in a single language.

(1) {N,NUM,NSG}+{N,NUM,OBL,SG}+{N,NUM,OBL,NSG}+
 {N,NUM,ACC,SG}/{N,NUM,SG}. This pattern of mixture is found in Modern Standard Arabic (Afro-Asiatic; Saudi Arabia and adjacent areas) and Berber (Ayt Seghrouchen Middle) (Afro-Asiatic; Morocco).

(6.32) Arabic (Modern Standard) (Holes 1995)

- a. {N,NUM,NSG}
 kita:b-a:ni θna:ni
 book-DU two.M
 'two books' (p. 173)

- b. {N,NUM,OBL,NSG}
 ?arba'at-u ?awla:d-in
 four.F-NOM boy.PL-GEN
 'four boys' (p. 173)

- c. {N,NUM,ACC,SG}
 xams-a 'ašarat-a bint-an
 five.M-ACC ten.F-ACC girl-ACC
 'fifteen girls' (lit. 'fifteen as regards girl') (p. 174)

- d. {N,NUM,OBL,SG}
mi ʔat-u bint-in
hundred-NOM girl-GEN
'a hundred girls' (p. 175)

(2) {N,NUM} and/or {N,NUM,NSG}+{N,NUM,CLF}+{N,NUM,CLF,NSG}: This pattern is found in Armenian (Eastern) (Indo-European; Armenia) and Persian (Indo-European; Iran), as shown in examples below. Note that it is highly possible that Armenian (Eastern) may also have {N,NUM,NSG} like Persian and most other Indo-European languages. However, we lack an example of that structure.

(6.33) *Persian (Mahootian 1997: 195)*

- a. {N,NUM,(CLF)}
bist (-jeld) ketab
twenty-(CLF) book
'twenty books'
- b. {N,NUM,(CLF),NSG}
do (-ta) pesæ-r-a
two (-CLF) boy-PL
'the two boys'

(3) {N,NUM}+{N,NUM,NSG}+{N,NUM,CLF}: This pattern is found in Buru (Austronesian; Indonesia) and Manchu (Altaic; China). Below are examples taken from Buru.

(6.34) *Buru (Grimes 1991)*

- a. {N,NUM,NSG}
huma-r polo
house-PL ten
'ten houses' (p. 295)

- b. {N,NUM,(CLF)}
- | | | |
|-------------|------------------|------------|
| <i>fafu</i> | (<i>kisen</i>) | <i>rua</i> |
| <i>pig</i> | (CLF) | <i>two</i> |
- 'two pigs' (p. 307)

(4) {N,NUM,NSG}+{N,NUM,NSG,DPM}: This pattern is found in Degema (Niger-Congo; Nigeria). The reader is referred to §5.3.6 for double plural marker (DPM).

(6.35) *Degema* (Kari 1997, 2004)

- a. {N,NUM,NSG}
- | | |
|---------------|---------------|
| <i>iyów</i> | <i>átuuw</i> |
| <i>twenty</i> | <i>hat.PL</i> |
- 'twenty hats' (1997: 36)

- b. {N,NUM,NSG,DPM}
- | | | |
|-----------------|------------|------------|
| <i>ímó</i> | <i>mé</i> | <i>va</i> |
| <i>child.PL</i> | <i>DPM</i> | <i>two</i> |
- 'two children' (2004: 209)

(5) {N,NUM}+{N,NUM,CLF,NSG}: This pattern is found in Halkomelem (Musqueam) (Salishan; Canada).

(6.36) *Halkomelem* (Musqueam) (Suttles 2004: 66)

- a. {N,NUM,CLF,NSG}
- | | |
|--------------------------------|-------------------|
| <i>t'xámə̀ə</i> | <i>stənténə̀y</i> |
| <i>six.CLF</i> (lit. 'person') | <i>woman.PL</i> |
- 'six women'

- b. {N,NUM}
- | | |
|--------------|----------------|
| <i>isə̀ə</i> | <i>mə̀lləs</i> |
| <i>two</i> | <i>raccoon</i> |
- 'two raccoons'

(6) {N,NUM,CLF}+{N,NUM,RMS}: This pattern is found in Kambera (Austronesian; Indonesia).

(6.37) *Kambera (Klamer 1998)*

a. {N,NUM,CLF}

<i>tailu</i>	<i>mbua</i>	<i>kajawa</i>
<i>three</i>	<i>CLF</i>	<i>papaya</i>

'three papayas' (p. 93)

b. {N,NUM,RMS}

<i>tau</i>	<i>ma-dua</i>
<i>person</i>	<i>RMS-two</i>

'two persons/people' (lit. 'people that are two') (p. 139)

(7) {N,NUM,NSG}+{N,NUM,CLF}: This pattern is found in Kolami (Dravidian; India).

(6.38) *Dravidian (Subrahmanyam 1998: 306)*

a. {N,NUM,CLF,NSG}

<i>pāj</i>	<i>jen</i>	<i>mās-ur</i>
<i>five</i>	<i>CLF.M</i>	<i>man-PL</i>

'five men'

b. {N,NUM,NSG}

<i>ayd</i>	<i>mās-ur</i>
<i>five</i>	<i>man-PL</i>

'five men'

(8) {N,NUM,CLF}+{N,NUM,CLF,NSG}+{N,NSG}: This pattern is found in Nez Perce (Penutian; Oregon, Washington) and Yagua (Peba-Yaguan; Peru).

(6.39) *Nez Perce* (Rude 1985)

a. {N,NSG}

ʔasqá-p-iin

brother-DU

'two brothers' (p. 77)

b. {N,NUM.CLF,NSG}

páax-loo' iwéepne-me

five-HUM wife-PL

'five wives' (p. 81)

c. {N,NUM,CLF}

le'épti-t wax náaqc wa'wáaʔam

twenty-NONHUM and one.NONHUM trout

'twenty one trout' (p. 81)

However, two languages are claimed to be lacking numerals proper; subsequently they do not have $CNNC_{NSG}$. Once again, Wari' and Pirahã illustrate this type (cf. §4.4), although at present, in Wari', Portuguese numerals have been borrowed.

The types just described above are treated as basic types in the languages and are also probably used as subsidiary ones. Besides, there are other possible structural patterns in addition to those just described. These structural patterns occur with low frequency and employed as the subsidiary modes in some languages.

Again, Imonda (Border; Papua New Guinea) shows a striking $CNNC_{NSG}$ —that is, the construction is composed of a noun, a numeral and a non-plural marker (or {N,NUM,NPL}). The non-plural marker has already been mentioned in §5.1.4.3, so it will not receive further treatment here. The example is illustrated again for convenience, though.

(6.40) *Imonda* (Seiler 1985: 39)

agō-ianèi sabla

woman.PL-NPL two

'two women'

In some numeral classifier languages, nouns might not be required in $CNNC_{NSG}$ if the numeral classifier has the same form as the head noun. This type may be coded as {NUM,CLF}, illustrated by Thai and Kuna (Chibchan; Colombia, Panama). Examples are given in (6.41) and (6.42) respectively.

(6.41) *Thai (own knowledge)*
sǎ:m khon
three CLF (lit. 'human')
'Three humans'

(6.42) *Kuna (Holmer 1946: 190)*
kwa-po
CLF-two
'two nuts' (NB: kwa kwa 'nut, nuts')

Moreover, there exists a language using a numeral marker for persons (PNUM) (cf. §5.3.5) in $CNNC_{NSG}$. This structural pattern is found in Maori (Austronesian; New Zealand).

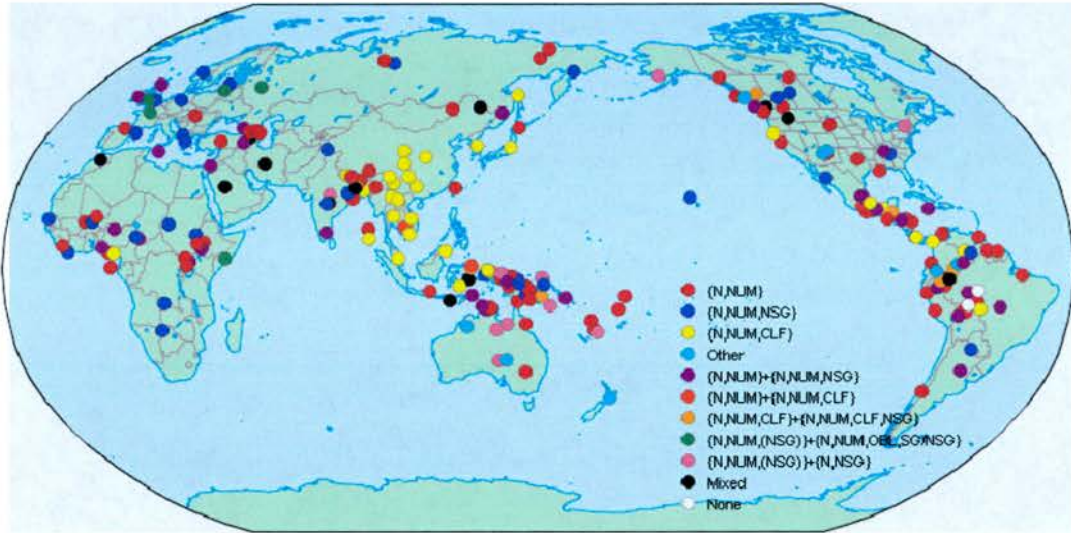
(6.43) *Maori (Bauer 1993: 496)*
{N,NUM,PNUM}
Toko-rima oona tuaakana
PNUM-five GEN.PL.3SG brother
'He had five older siblings'

Finally, although the examples below may not count in $CNNC_{NSG}$ at all, they illustrate another possibility of quantificational expression without numerals. In Lavukaleve (Solomons East Papuan; Solomon Islands), there are a few nouns which refer to a group of ten. These nouns denote culturally important things. Note that the pairs of singular forms and their counterparts do not seem to share the same root.

<i>foe</i>	<i>'pig'</i>	<i>kolo</i>	<i>'ten pigs'</i>
<i>mita'keu</i>	<i>'dog'</i>	<i>feil</i>	<i>'ten dogs'</i>
<i>fo'sal</i>	<i>'fish'</i>	<i>lolu</i>	<i>'ten fish'</i>
<i>urio</i>	<i>'coconut crab'</i>	<i>legom</i>	<i>'ten coconut crabs'</i>
<i>karu</i>	<i>'possum'</i>	<i>koku</i>	<i>'ten possums'</i>

6.4 Geographical distribution of $CNNC_{NSG}$

This section is concerned with the patterns of occurrence of language types with respect to $CNNC_{NSG}$ in the world's languages, as depicted in Map 6.4 below.



Map 6.4 Geographical distribution of $CNNC_{NSG}$



Map 6.5 New Guinea and Northern Australia (zoomed)

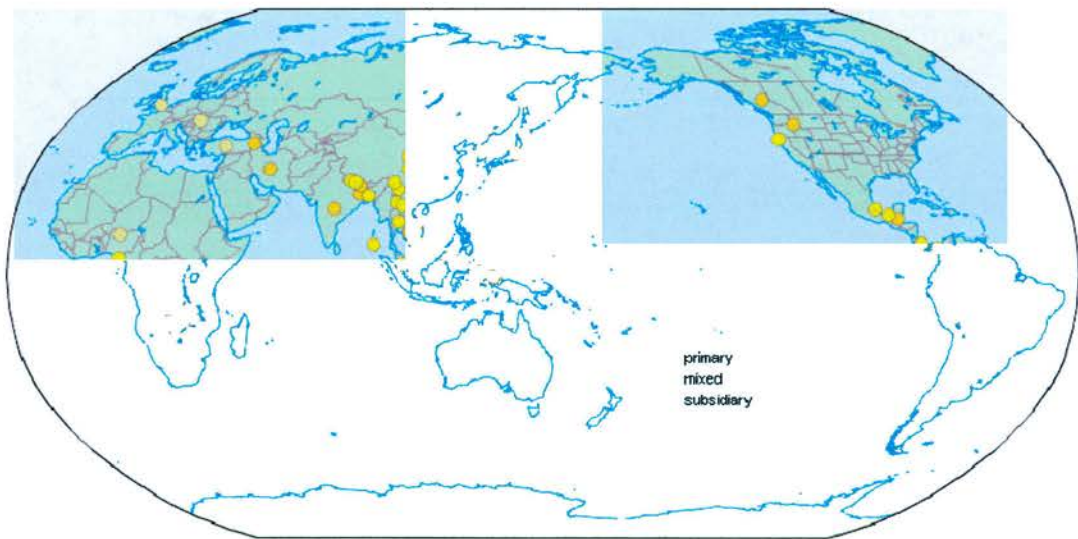


Map 6.6 Mesoamerica (zoomed)

Map 6.4 shows that the world's languages can basically be divided into 4 major groups, namely $\{N,NUM\}$, $\{N,NUM,NSG\}$, $\{N,NUM,CLF\}$ and the mixture of $\{N,NUM\}$ and $\{N,NUM,NSG\}$ with other possibilities being more minor. The first and most common group is $\{N,NUM\}$ (shown in red). This type spreads through most major

regions except Western Europe, Southeast Asia and (parts of) East Asia. The second most common is {N,NUM,NSG} (shown in blue). This type is fairly common in many major regions of the world, typically Western Europe and Africa but totally absent in Southeast Asia. Although this type as a primary mode seems to be relatively rare in Australia and South America, this type is in fact present along with {N,NUM} and {N,NSG} in some languages of the two areas as Mixed, shown in purple ({N,NUM,NSG}+{N,NUM}) and pink ({N,NUM,NSG}+{N,NSG}).

As for {N,NUM,CLF}, the distribution is not different from what is already described in CNNC_{SG} (cf. § 6.1), that is to say, they are present around the Pacific Rim. This section, however, will discuss a bit further the modes of occurrence with various degrees of obligatoriness of {N,NUM,CLF} in the world's languages. Map 6.7 below shows the geographical distribution of {N,NUM,CLF} in the world's languages with respect to degree of obligatoriness (based on the current sample). The map also shows three modes of occurrence of {N,NUM,CLF}, namely *primary*, *mixed* (those which are employed equally with other types) and *subsidiary*, represented by the yellow, gold, and khaki colours respectively.



Map 6.7 Geographical distribution of {N,NUM,CLF} with respect to degrees of obligatoriness

As shown in Map 6.7 above, the degree of obligatoriness and the frequency of {N,NUM,CLF} are less in Western Asia and eastern Europe and absent completely in Western Europe. English shows a numeral classifier-like word (e.g. *four hundred head of cattle*). However, the instance is exceptional, and therefore English is not regarded as a numeral classifier language. In Western Asia and eastern Europe, {N,NUM,CLF} is never used as the single primary means of forming CNNCs, as evidenced by Persian (Indo-European), Eastern Armenian (Indo-European), Hungarian (Uralic) and Turkish (Altaic). Interestingly, a small pocket of numeral classifiers used obligatorily is also present in western Africa (Aikhenvald 2000: 121-123; Gil 2005: 226-229; Ikoru 1994). One instance from Kana (Niger-Congo; Nigeria; Ikoru 1994) is observed in the current sample. However, apart from the regions mentioned, {N,NUM,CLF} is completely absent.

The findings described above accord well with previous studies of the distribution of numeral classifiers (for example, Nichols 1992: 132-133, Aikhenvald 2000: 121-123, and Gil 2005: 226-229). Still, it has not yet been made clear whether numeral classifiers across languages have developed in isolation or have developed in only one region and have then spread to other regions. However, based on the geographical distribution illustrated, it seems likely that numeral classifier systems might originate by polygenesis. It is possible that at least, the classifier systems may have originated and developed individually in three different places, namely Southeast Asia, the Americas and Africa, due to the fact that the three hotbeds of numeral classifier systems are very far from each other. As for the southern part of Southeast Asia, Oceania and Western Asia, due to the fact that the use of numeral classifiers is optional, it is possible that the numeral classifier systems have developed there due to language contact.

The next type is a group of languages classified as *Other* (shown in light blue). Structural patterns of CNNC_{NSG} belonging to this group are represented by a limited number of languages. The distribution of this type is also interesting, as it is coincidentally mainly found in the New Worlds like Australia-New Guinea and the Americas.

Regarding the languages showing mixtures, there are two regular distributions. The first is the mixture of {N,NUM} + {N,NUM,NSG} (shown in purple). The cause of mixture is quite interesting. It may be noticed that in many places this pattern is surrounded by languages with {N,NUM} and languages with {N,NUM,NSG}, for example, Nuuchahnulth (Wakashan; Canada), Leti (Austronesian; Indonesia), and Maale (Afro-Asiatic; Ethiopia). However, no evidence indicates that the mixture of the patterns is due to language contact. It seems more likely that this pattern may develop in isolation due to internal variation of CNNC_{NSG} in the language; that is to say, the use of non-singular markers varies to the animacy hierarchy, typically {N,NUM} for non-human nouns or inanimate nouns and {N,NUM,NSG} for human nouns or animate nouns. For example, in Hausa (Afro-Asiatic; Niger, Nigeria), {N,NUM} is used with inanimate nouns, while {N,NUM,NSG} is used with animate nouns (Schuh 1999: 1). In addition, the use of non-singular markers may be subject to the value of numerals modifying nouns. For example, in Hebrew (Afro-Asiatic; Israel), {N,NUM,NSG} is used the numerals 3-10, whereas {N,NUM} is used with the numerals greater than 10 (Ojeda 1994: 1)

The other systematic pattern of a mixture is {N,NUM,CLF}+{N,NUM,CLF,NSG} (shown in orange). This group of languages is quite interesting because normally numeral classifiers tend not to co-occur with plural markers (cf. Sanches and Slobin 1973). The map shows a clear distribution such that this pattern occurs in the areas that the languages with plural markers have been in contact with numeral classifiers. The areas where the languages of this type are mostly found are the Americas and Papua New Guinea where the plural marking languages and numeral classifier languages are in contact. An instance of this type is also found in Malto (Dravidian; India), a plural marking language. The language has presumably adopted numeral classifiers from Indo-Aryan languages in India (cf. Emeneau (1956: 14). The structural pattern of {N,NUM,CLF,NSG} can be observed in some other non-typical areas of numeral classifier languages as well but as a subsidiary mode or mixed with

other structural patterns of $CNNC_{NSG}$, such as in Korean (Altaic; Korea), Armenian (Indo-European; Armenia) and Persian (Indo-European; Iran).

Another mixed type is $\{N,NUM,(NSG)\}+\{N,NUM,OBL,SG/NSG\}$ (shown in green). This type of mixture is represented by a limited number of languages, four of which are found in northern Europe, i.e. Russian (Indo-European; Russia), Welsh (Indo-European; Wales), French (Indo-European; France) and Lithuanian (Indo-European; Lithuania). One instance is also observed in Somali (Afro-Asiatic; Somalia). This type of mixture seems to be a remnant of Indo-European counting expressions, reflecting that in this family these numerals may belong to different word classes originally. That is to say, high round numerals are more noun-like, and require an oblique case in $CNNC_{NSG}$, whereas the low-valued numerals are more adjective, and can modify nouns directly (for further discussion, see §9.6).

The next type is the mixture of $\{N,NUM,(NSG)\}+\{N,NSG\}$ (shown in pink). This type is commonly found in Australia-New Guinea where $\{N,NSG\}$ is notable. This might be due to the fact that numerals are rarely used in this region (Dixon 2002: 67) but rather the non-singular numbers are used instead, typically dual and trial for the low-valued numbers. As for the languages classified as *mixed* (shown in black), there is a variety of mixtures and there is no clear distribution to the pattern.

Finally, like $CNNC_{SG}$, two languages, in the current sample, namely Wari' and Pirahã, spoken in Brazil, South America are reported as having no numerals proper and hence lacking $CNNC_{NSG}$.

6.5 Conclusion

In this chapter, the range of possible types of CNNCs present in the world's languages has been illustrated along with their frequencies and geographical distribution. The structural types which are subsidiary modes have not been taken into consideration when categorizing types of languages, but they have been described. For $CNNC_{SG}$, there are basically two major types with very few other possibilities. They are $\{N,NUM\}$, the most common language type occupying almost all the major regions of the world, and $\{N,NUM,CLF\}$ which is fairly common around the Pacific Rim. Regarding $CNNC_{NSG}$, there are basically four major types, namely $\{N,NUM\}$, $\{N,NUM,NSG\}$, $\{N,NUM,CLF\}$ and the mixture of $\{N,NUM\}$ and $\{N,NUM,NSG\}$, with some other possibilities. The establishment of these types is necessary for an investigation into the evolution of CNNCs.

7 CNNCs and Noun Class: An Implicational Universal?

In Chapter 6, various language types of CNNCs were established. In Chapter 7, we will move further to see to what extent we can generalize systematic patterns of occurrence in these language types. This step is referred to in Croft (2003:1) as *typological generalisation*. One aspect of typological generalisation uses *implicational universals*, which state that the presence of one linguistic feature implies the presence of another linguistic feature (Croft 2003: 1). An example is Greenberg's universal 34 (1963:74): “*If a language has the category of gender, it always has the category of number.*” This implicational universal can be paraphrased by stating that the presence of the category of gender in a language implies the presence of the category of number.

One of the grammatical systems potentially related to CNNCs is *classifier systems* (cf. §5.2). It is generally recognized that classifier systems associate with *number systems* in several aspects. This can be illustrated by Greenberg's universal 34 as mentioned above, and also by the implicational universal proposed by Sanches and Slobin (1973:4), which states that the presence of number marking on nouns and the presence of numeral classifiers are likely to be in complementary distribution (cf. §2.2.1). Interdependencies between classifier systems and number systems, especially at the *morphological level*, are widely reported in the literature (see, for example, Aikhenvald 2000: 242-252).

In this chapter, hypothesised generalisations are made to the distribution of $CNNC_{NSG}$. Based on a superficial observation, it seems likely that $CNNC_{NSG}$ occurs systematically across languages in association with one type of classifier systems, namely *noun class* (NC) (hereafter also including *gender system*, cf. §5.2.3.3).

¹⁶ It is hypothesised that a language where $\{N, NUM, NSG, (X)\}$ ¹⁷ is employed,

¹⁶Note that *gender* or *noun class* here deals with nouns only. Gender distinction in personal pronouns, such as English *he* and *she*, is not taken into consideration. English is therefore treated here as a language without noun class or gender.

either as a primary mode or a non-primary (i.e. mixed or subsidiary) mode of $CNNC_{NSG}$, tends to have the category of noun class. An example is French which is a noun class language and has $\{N,NUM,NSG\}$ and $\{N,NUM,OBL,NSG\}$. On the other hand, it is also hypothesised that a language in which $\{N,NUM,NSG,(X)\}$ is not present at all tends to lack the category of noun class (even if the language has the category of number). One such language of this type is Hungarian which is a non-noun class language and does not have $\{N,NUM,NSG,(X)\}$. The two related hypotheses imply that if one finds a language possessing noun class systems, one should expect $\{N,NUM,NSG,(X)\}$ to be used in the language, at least as a non-primary mode of $CNNC_{NSG}$. Instead, if one finds a language without noun class systems, one should expect any type, such as $\{N,NUM\}$, $\{N,NUM,CLF\}$, $\{N,NSG\}$ and so on, but not $\{N,NUM,NSG,(X)\}$ in the language. The hypotheses are posited on the assumption that the language in question has $CNNC_{NSG}$. The languages with no $CNNC_{NSG}$, namely Wari' and Pirahã are excluded from these hypotheses, whether or not they are reported to have noun class systems.

The chapter is structured as follows: In §7.1, the motivation underlying these hypotheses is discussed. Since the typological generalisations being made are hypothetical by nature, a tool for determining whether the hypotheses are *statistically significant* is then required—that is, a means of ascertaining whether or not the generalisations are likely to have arisen by chance. The model which is employed here to test the hypothesised implicational universal is one designed by Dryer (2003). The model for testing statistical significance is reviewed in §7.2. In §7.3, the hypothesised generalisations are tested. Finally, the last section (§7.4) an explanation for the relationship between $\{N,NUM,NSG,(X)\}$ and noun class systems is given.

¹⁷ The notation $\{N,NUM,NSG,(X)\}$ includes structural types of $CNNC_{NSG}$ consisting of at least the *three* constituents, namely N, NUM and NSG; X is any quantificational extra element, such as an oblique marker, or a numeral classifier (cf. §5.1-§5.3). Hence, $\{N,NUM,NSG\}$, $\{N,NUM,CLF,NSG\}$, $\{N,NUM,OBL,NSG\}$ and $\{N,NUM,NSG,DPM\}$ are instances of $\{N,NUM,NSG,(X)\}$, whereas $\{N,NUM\}$, $\{N,NSG\}$, and $\{N,NUM,CLF\}$ are not.

7.1 What motivates the hypotheses?

The two hypotheses are motivated by a superficial observation on the current data that in the predominantly noun class areas, typically Africa and Europe (cf. Corbett 2005: 126-129), we often see {N,NUM,NSG,(X)}, either as a primary mode or a non-primary mode of $CNNC_{NSG}$. On the contrary, in the areas where noun class is not strong or relatively rare, such as western and north eastern Asia, Southeast Asia & Oceania (cf. Corbett 2005: 126-129), {N,NUM,NSG,(X)} seems to be less observed as well, even in languages with number marking systems such as Hungarian. Refer to Map 6.4 and Table B in Appendix 1 for the geographical distribution of {N,NUM,NSG,(X)}. Therefore, it is interesting to investigate further whether the co-incidence as observed is attributable to typological, genetic or areal phenomena. If the observed co-incidence is typological, it is supposed to occur in languages regardless of genetic or areal relationship, in other words, the relationship between the two features is not restricted to particular genetic groups of languages or geographical areas.

This observation also leads to a consideration of Greenberg's universal 34 (1963:74): "*If a language has the category of gender, it always has the category of number.*" Considering Greenberg's universal 34 again, one may wonder whether or not it may also be true to frame it the other way round. That is, is it necessary that a language with the category of number always has the category of gender? Based on the WAL database, the answer is 'no'. Below is the information on the correlation between the category of number and the category of gender generated with the *Interactive Reference Tool* (developed by Hans-Jörg Bibiko 2005) of *The World Atlas of Language Structures* (Haspelmath, Dryer, Gil and Comrie 2005). The language sample of 119 languages is considered in association with *the occurrence of nominal plurality* (Haspelmath 2005) and *number of genders* (Corbett 2005). The presence of the two features is distinguished by *present* (+) and *absent* (-). When these two dimensions (namely the two grammatical features and the presence of the grammatical two features) are combined, we get 4 values shown in Table 7.1 below. Refer to Appendix 2.4 for a list of languages from Table 7.1.

Features	[no. of languages, genera, families]
+Plural/+Gender	[54:37:20]
+Plural/-Gender	[55:48:31]
-Plural/-Gender	[7:7:6]
-Plural /+Gender	[3: 3: 2]
Total	[119:95:59]

Table 7.1 The relationship between plural and gender

According to the information given in WALS, among the 57 languages with + Gender, 54 languages also have the category of number but only 3 languages do not. When considering the 109 languages with +Plural, about 54 languages have gender and 55 languages do not. This means that half of the languages which have number do not possess the category of gender. So, overall, based on this information, we may add the phrase “but not necessarily vice versa” to Greenberg’s Universal 34, so that it reads, “If a language has the category of gender, it always has the category of number *but not necessarily vice versa*”.

This new statement implies that the languages which possess the category of number can be divided into 2 groups—namely, one with noun classes and the other without. In accordance with the hypotheses proposed above, it might be generalised that the number-marking languages with noun classes are those in which the non-singular marker is likely to be employed in forming $CNNC_{NSG}$. On the other hand, in those number-marking languages without noun classes, the non-singular marker tends not to play a part in $CNNC_{NSG}$. This can be illustrated as in Figure 7.1 below.

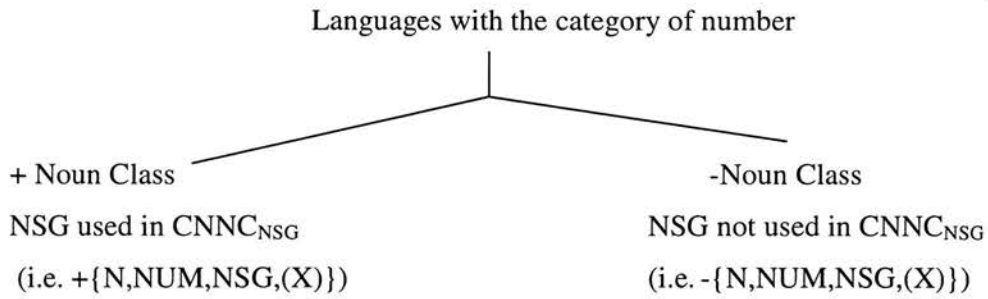


Figure 7.1 Two types of languages with the category of number

However, the hypothesized generalizations discussed above are speculative. The significance of the correlation between the presence of $\{N,NUM,NSG,(X)\}$ and the presence of noun class will therefore be tested. To see whether or not the typological generalisations are statistically significant, a statistical test employed to determine whether the generalisations proposed are statistically significant is reviewed in the next section.

7.2 How can we test an implicational universal?

According to Dryer (2003), the commonly used tests for statistical significance, such as Fisher's Exact Test and the Chi-Square Test, do not suit most typological generalisations because the tests require the instances in the sample to be independent. Most observations in typological samples are however likely to be dependent on each other in one way or another—they are often genetically, geographically or typologically related, for example. Therefore, any statistical test which is used should be designed specifically for typological samples. One method for testing the statistical significance of implicational universals is proposed by Dryer (2003). Here we review the relevant parts of this method.

Implicational universals are always in the logical form of 'if P then Q', which means that if a language has property P, then it will also have property Q. For example, Greenberg's universal 34, "If a language has the category of gender, it always has the category of number," can be written in a simple logical form as "if *gender* then *number*". So, from the outset, it will be easier if the hypothesised generalisation is written in logical form as *if P, then Q*.

Dryer (2003: 110-111) proposes that the generalisation will be justified (i.e. statistically significant) if it is true in *all* six geographical areas (i.e. not restricted to particular areas), namely Africa (AFR), Eurasia (EUR), Southeast Asia & Oceania (SEA&OCE), Australia-New Guinea (AUS-NEW), North America (NAM), and South America (SAM). This is because there is only one chance in 64 (i.e. $1/2^6$) of this phenomenon occurring. The logic behind the figure $1/64$ or $1/2^6$ comes from the basic assumption proposed by Dryer (1989a) that the world's languages are divided into 6 geographical areas. The chance that one area will show a given preference is $1/2$. Then, the chance that 2 areas both will show a given preference must be half less, i.e. $1/2 \times 1/2$, yielding $1/2^2$ (i.e. $1/4$). Then the chance that 3 areas all will show a given preference must be $1/2 \times 1/2^2$, yielding $1/2^3$ (i.e. $1/8$). So then the chance that 4, 5, and 6 areas all

will have show a given preference must be less and less, finally the chance that six areas will show a given preference will be reduced to $1/2^6$ (i.e. $1/64$).

To test the statistical significance of the generalisation ‘if P, then Q’ we then will compare the supporting evidence with the counter-evidence for the generalisation in tabular form, as shown in the sample in Table 7.2. The figures in the table represent the number of genera containing languages of the given type in the six geographical areas. Note that hereafter the figures in the cells represent the number of genera. The number of genera is used instead of the number of languages to reduce the genetic bias. The figures in bold represent the larger number, suggesting that the property in question is the more frequent type in that area.

	AFR	EUR	SEA&OCE	AUS- NEW	NAM	SAM	Total
P&Q	5	7	8	8	7	9	44
P& -Q	4	6	0	4	3	1	18

Table 7.2 Property Q in languages with property P

In Table 7.2, among the languages with property P, property Q outnumbers property non-Q in all six areas—in other words, the supporting evidence for the implicational universal *If P then Q* is greater than the counter-evidence in all six areas. For example, in AFR, 5 genera with property P have property Q, whereas 4 genera with property P do not have property Q. So, the hypothesised implicational universal is probabilistically valid in AFR. Considering the remaining areas in the same way, it is found that the overall outcome for each area is the same as that obtained for AFR. Therefore, it can be concluded that the hypothesised implicational universal expresses a valid probabilistic generalisation.

However, based on the data in Table 7.2, we do not know exactly whether or not the languages which are not P (or $\neg P$) are likely to have Q as well and whether or not the tendency of “if $\neg P$ then Q” is stronger than the case of “if P then Q”. Therefore, Dryer

(2003: 111) posits two degrees of the validity of a typological generalisation or an implicational universal, namely, *weak* and *strong* as follows:

“If P then Q” is true *in a weak sense* if languages with property P more often have property Q regardless of whether languages with property not-P also tend to have property Q. (italics original)

“If P then Q” is true *in a strong sense* if languages with property P more often have property Q and the tendency for languages with property P to have property Q is significantly stronger than the tendency for languages with property not-P to have property Q. (italics original)

Based on Table 7.2, at this stage the hypothesised generalisation can be said to be true in a weak sense. Therefore, we need to look at data for languages with -P to determine whether the generalisation might also be true in a strong sense. See Table 7.3.

	AFR	EUR	SEA&OCE	AUS- NEW	NAM	SAM	Total
-P&Q	3	2	1	6	5	4	21
-P&-Q	11	19	17	14	9	11	81

Table 7.3 Property Q in languages with property -P

Table 7.3 demonstrates that the languages with property -P are unlikely to have property Q. The next step is to test the hypothesis that the preference for property Q among languages with property P is stronger than the same preference in languages which are -P. If it is, the hypothesis will be true in a strong sense. We can compare the proportions of genera in each of the six areas, as in Table 7.4. The proportion is computed by taking the figure in a cell in row 1 and dividing that by the sum of the same number plus the figure in the corresponding cell in row 2. For example, in Table 7.4, the proportion of -P & Q in Africa is calculated by taking 3 divided by the sum of 3+11 (i.e.14), yielding 0.21. As for the rightmost column, titled ‘*mean*’, the cell shows the number of occurrences of the property Q averaged over each area. The result is

calculated by taking the sum of all cells in the row and then dividing the sum by 6 (i.e. 6 areas). The overall result is shown in Table 7.4.

	AFR	EUR	SEA&OCE	AUS-NEW	NAM	SAM	Mean
P	0.56	0.54	1	0.67	0.70	0.90	0.73
-P	0.21	0.10	0.10	0.30	0.36	0.27	0.22

Table 7.4 Proportions of genera with property Q among languages with P and -P

Table 7.4 shows that the languages with P show a stronger preference for Q than do those with -P in all six areas. Since the chance of all six areas reflecting the same pattern is only 1 in 64, it can be concluded that languages with P show a significantly stronger preference for Q than do languages without P. Hence, we can say that the hypothesised typological generalisation is true in a strong sense.

In practice, however, the situation may not be as neat as illustrated above—one area may not conform to the generalisation, as shown in Table 7.5 below (adapted from Dryer (2003: 113)).

	AFR	EUR	SEA&OCE	AUS-NEW	NAM	SAM	Total
P&Q	7	1	8	1	16	4	37
P&-Q	0	0	0	2	3	1	6

Table 7.5 Property Q in languages with property P

According to Table 7.5, property Q is more common among the languages with property P in 5 out of the 6 areas and the preference for property Q in these areas is quite strong in the 5 areas. However, Australia-New Guinea does not conform to the generalisation “if P then Q”. The preference for property -Q among the languages with property P in this area is more common by 2 genera to 1. However, since the preference for property Q in the other five areas is quite strong and the difference of the preference

for proper Q in Australia-New Guinea is very small, Dryer (2003: 113) holds that the data in Table 7.5 show evidence for a real linguistic preference, in other words, the generalisation “if P then Q” is true (at least in a weak sense).

As a rule of thumb, I adopt the practice of tentatively accepting a pattern as reflecting a real linguistic preference if a type is more common in 5 out of the 6 areas, if the preference for that type is quite *strong* in those other 5 areas, and if the greater number of genera in the one exceptional area is by a *relatively small margin*. (Dryer 2003: 113, italics mine)¹⁸

Dryer (2003) also shows an example of data of which a real linguistic preference fails to satisfy the test, as shown in Table 7.6 below (adapted from Dryer (2003: 113)).

	AFR	EUR	SEA&OCE	AUS- NEW	NAM	SAM	Total
P&Q	6	9	14	7	22	8	66
P&-Q	31	0	12	6	0	0	49

Table 7.6 Property Q in languages with property P

According to Table 7.6, among languages with property P, property Q is common than property -Q in 5 out of the 6 areas. In one area, namely Africa, however, property -Q is more common. It is also noticeable that the preference for property -Q in Africa is by far more common (i.e. 31 genera to 6), so the preference for property -Q is quite strong. Moreover, among the 5 areas, 2 areas, namely Southeast Asia & Oceania and Australia-New Guinea, the preferences for Q and -Q in the 2 areas are very slightly different. In this case, Dryer (2003: 113) does not hold that the data such as Table 7.6 reflect a real linguistic preference.

Moreover, Dryer (2003) suggests that there can be a significant dependency between P and Q that do not involve significant implicational universals. This situation can be illustrated by Table 7.7 below (adapted from Dryer (2003: 117)).

¹⁸ It is not clear from the source, however, how to measure the difference between ‘strong’ and ‘not strong’ or ‘relatively small margin’ or ‘not relatively small margin’. This might be a problem in a situation where one area does not conform to the test.

	AFR	EUR	SEA&OCE	AUS-NEW	NAM	SAM	Total
P&Q	7	6	12	5	20	5	55
P&-Q	17	0	9	2	4	0	32
-P&Q	2	7	2	9	4	4	28
-P&-Q	10	3	6	7	7	1	44

Table 7.7 Property Q and Property P

According to Table 7.7, the first two rows suggest that while property Q is more common than $\neg Q$ in 5 out of the 6 areas. Africa is the only one area fails to satisfy the test. In Africa, property $\neg Q$ is more common by 17 genera to 7. Therefore, the hypothesised implicational universal “if P then Q” is not statistically significant. In relation to the last two rows, showing the opposite trend in languages with property $\neg P$, the table shows that the preference for property $\neg Q$ in languages with $\neg P$ is valid only in 4 areas. Two areas, namely Africa and Southeast Asia & Oceania, fail to conform to the generalisation “if $\neg P$ then $\neg Q$ ”. Hence, the implicational universal is not valid.

According to Table 7.7, at this stage we know that there is no a significant implicational universal “If P then Q”. However, comparing proportions of genera that have property Q among languages with property P and $\neg P$, as shown in Table 7.8, it turns out that there is “a statistically significant dependency” between property P and property Q in all areas “without there being a significant implicational universal” (Dryer 2003: 118). In Africa, for example, although the evidence does not support the hypothesised implicational universal “if P then Q”, the proportions of genera that are Q in languages with P are higher than those with $\neg P$.

	AFR	EUR	SEA&OCE	AUS- NEW	NAM	SAM	Mean
P	0.29	1.00	.57	.71	.83	1.00	.73
-P	.17	.70	.25	.35	.36	.80	.32

Table 7.8 Proportions of genera that are Q among languages with P and -P

7.3 Testing the hypothesised implicational universals

The hypothesised implicational universals proposed in this chapter is that “a language in which {N,NUM,NSG,(X)} is employed, either as a primary mode or a non-primary mode of CNNC_{NSG}, tends to have the category of noun class, whereas a language in which {N,NUM,NSG,(X)} is not employed at all tends to lack the category of noun class, (even if the language has the category of number).” The two generalisations can be written in a form of “if P then Q” as “if +{N,NUM,NSG,(X)} then [+NC]” and “if -{N,NUM,NSG,(X)} then [-NC]”.

The evidence in favour of the first generalisation (i.e. if +{N,NUM,NSG,(X)} then [+NC]) is presented in Table 7.9. The lists of languages and genera of each cell in the top row and the second row are provided in Appendix 2.5.1 and Appendix 2.5.2 respectively.

	AFR	EUR	SEA&OCE	AUS- NEW	NAM	SAM	Total
+{N,NUM,NSG,(X)}& [+NC]	11	16	0	9	3	6	45
+{N,NUM,NSG,(X)}& [-NC]	5	6	3	3	10	4	31

Table 7.9 Property [+NC] in languages with + {N,NUM,NSG,(X)}

Table 7.9 shows that among the languages which have {N,NUM,NSG,(X)}, the property [+NC] is more common than the property [-NC] in 4 out of the 6 geographical areas, namely Africa, Eurasia, Australia-New Guinea and South America. This means that the languages of {N,NUM,NSG,(X)} in those areas tend to have the category of noun class. It can be noticed that the preference for [+NC] among the languages of {N,NUM,NSG,(X)} is relatively strong in noun class areas, i.e. Africa, Eurasia (Europe) and Australia-New Guinea.

Two areas do not conform to the generalisation “if $\{N,NUM,NSG,(X)\}$ then $[+NC]$ ”. In Southeast Asia & Oceania, among the languages which we have information about noun classes, 3 genera are reported to have $\{N,NUM,NSG,(X)\}$, and the 3 genera prefer the property $[-NC]$. Hence, the property $[-NC]$ is more common by 3 genera to 0. In North America, it is found that the languages of $\{N,NUM,NSG,(X)\}$ prefer to have the property $[-NC]$, that is to say, the property $[-NC]$ is more common by 10 genera to 3. It can be said that the preference of $[-NC]$ over $[+NC]$ among languages with $\{N,NUM,NSG,(X)\}$ is quite strong in North America. If we hold that a generalisation which is statistically significant should be independently supported in all six areas or at least in five areas, the evidence in Table 7.9 therefore falls short of statistical significance for the generalisation “if $\{N,NUM,NSG,(X)\}$ then $[+NC]$ ”, since there are only 4 areas satisfying the test.

However, it is interesting to investigate further if there is a dependency without there being a significant implicational universal. So, we will first examine whether or not there is a preference for the property $[+NC]$ among the languages with $\{-N,NUM,NSG,(X)\}$. The result is shown in Table 7.10 below. The lists of languages and genera of each cell in the top row and the bottom row are provided in Appendix 2.5.3 and Appendix 2.5.4 respectively.

	AFR	EUR	SEA&OCE	AUS- NEW	NAM	SAM	Total
$\{-N,NUM,NSG,(X)\}$ & $[+NC]$	1	1	1	4	1	0	8
$\{-N,NUM,NSG,(X)\}$ & $[-NC]$	5	13	20	9	9	14	70

Table 7.10 Property $[+NC]$ in languages with $\{-N,NUM,NSG,(X)\}$

From Table 7.10, it can be seen that languages which are not $\{N,NUM,NSG,(X)\}$ do not prefer the category of $[+NC]$, rather they prefer the property $[-NC]$. This is true in all 6 geographical areas. The next step is therefore to test the

hypothesis that the preference for the property [+NC] among the languages with $\{N,NUM,NSG,(X)\}$ is stronger than the same preference in $\{-N,NUM,NSG,(X)\}$. If it is, the result would reflect that there is a *dependency* between $\{N,NUM,NSG,(X)\}$ and [+NC]. The result is shown in Table 7.11.

	AFR	EUR	SEA&OCE	AUS- NEW	NAM	SAM	Mean
$\{N,NUM,NSG,(X)\}$	0.69	0.32	0	0.75	0.23	0.60	0.43
$\{-N,NUM,NSG,(X)\}$	0.17	0.07	0.05	0.31	0.10	0	0.12

Table 7.11 Proportion of genera which have the property [+NC]

Table 7.11 shows that the proportions of genera which have the property [+NC] is higher among $\{N,NUM,NSG,(X)\}$ than among $\{-N,NUM,NSG,(X)\}$ in 5 out of the 6 areas. One area (SEA&OCE) does not conform to the generalization. Although the preference for [+NC] shows a relatively small difference by 0.05 to 0, it is hard to conclude that there is a statistically significant dependency between $\{N,NUM,NSG,(X)\}$ and [+NC]. According to Dryer (2003: 113), as a rule of thumb, if the tendency is quite strong at least in 5 areas, and the exceptional area shows a small difference, it is acceptable that there is a real linguistic preference. Although some areas, namely Africa and Eurasia appear to show a relatively strong tendency, 2 areas are hard to conclude that they also show a strong tendency, namely New Guinea (0.75 to 0.31) and North America (0.23 to 0.10). Therefore, there is no dependency between $\{N,NUM,NSG,(X)\}$ and [+NC]. The typological observation on the relationship between the two features should be attributable to geographical areas and genetic relationship rather than typology.

Next, we will test the second generalisation that a language in which $\{N,NUM,NSG,(X)\}$ is not present at all tends to lack the category of noun class. The generalisation can be written in a form of “if P then Q” as “if $\{-N,NUM,NSG,(X)\}$ then [-NC]”. The evidence in favour of this generalisation (i.e. if $\{-N,NUM,NSG,(X)\}$ then

[-NC]) is presented in Table 7.12. The lists of languages and genera of each cell in the top row and the bottom row are provided in Appendix 2.5.4 and Appendix 2.5.3 respectively.

	AFR	EUR	SEA&OCE	AUS-NEW	NAM	SAM	Total
-{N,NUM,NSG,(X)} & [-NC]	5	13	20	9	9	14	70
-{N,NUM,NSG,(X)} & [+NC]	1	1	1	4	1	0	8

Table 7.12 Languages with $\{-N,NUM,NSG,(X)\}$ and the property [+NC]

Table 7.12 shows that among the languages with the property $\{-N,NUM,NSG,(X)\}$, the property [-NC] is more common than the property [+NC] in all 6 areas. For example, in Africa, the languages of $\{-N,NUM,NSG,(X)\}$ having the property [-NC] are more common than those having [+NC] by 5 genera to 1. This means that in African languages which do not have $\{N,NUM,NSG,(X)\}$, we would expect them to be non-noun class languages rather than noun class languages. The table also shows that the tendency for the preference seems to be very strong in most areas. According to Dryer (2003)'s method, if the generalisation is valid in 6 areas, the phenomenon will be regarded as there being a real linguistic preference. The result shows that languages with the property $\{-N,NUM,NSG,(X)\}$ more often have the property [-NC]. Still, we do not know whether languages with the property $\{+N,NUM,NSG,(X)\}$ also tend to have the property [-NC]. Therefore, at this stage the hypothesised generalisation "if $\{-N,NUM,NSG,(X)\}$ then [-NC]" is true but only in a weak sense. Next, we will examine the opposite trend, that is, whether or not there is a preference for the property [-NC] among the languages with the property $\{+N,NUM,NSG,(X)\}$. The result is depicted in Table 7.13. The lists of languages and genera of each cell in the top row and the bottom row are provided in Appendix 2.5.2 and Appendix 2.5.1 respectively.

	AFR	EUR	SEA&OCE	AUS- NEW	NAM	SAM	Total
+{N,NUM,NSG,(X)}& [-NC]	5	6	3	3	10	4	31
+{N,NUM,NSG,(X)}& [+NC]	11	16	0	9	3	6	45

Table 7.13 Languages with +{N,NUM,NSG,(X)} and the property [-NC]

Table 7.13 shows that the languages with the property +{N,NUM,NSG,(X)} do not prefer the category of [-NC]. There are only 2 out of the 6 areas showing the preference for [-NC], namely Southeast Asia & Oceania (by 3 genera to 0) and North America (by 10 genera to 3), whereas there are 4 areas showing the preference for the category of [+NC]. For example, in Africa, among the languages with the property +{N,NUM,NSG,(X)}, the property [+NC] outnumbers the property [-NC] by 11 genera to 5. This means that in any African language, if {N,NUM,NSG,(X)} exists, we should expect the language to be a noun class language rather than non-a noun class language. Therefore, Table 7.13 shows that languages which have the property +{N,NUM,NSG,(X)} are unlikely to have property [-NC].

After we have seen the preference for [-NC] among the languages with -{N,NUM,NSG,(X)} and +{N,NUM,NSG,(X)}, the next step is to test the hypothesis that the preference for [-NC] among languages with -{N,NUM,NSG,(X)} is stronger than the same preference in languages with + {N,NUM,NSG,(X)}. If it is, the hypothesised generalisation will be true in a strong sense. We will now compare the proportions of genera which have the property [-NC] between the two types of languages in each of the 6 areas, as in Table 7.11 below.

	AFR	EUR	SEA&OCE	AUS- NEW	NAM	SAM	Mean
-{N,NUM,NSG,(X)}	0.80	0.93	0.95	0.69	0.90	1	0.88
+{N,NUM,NSG,(X)}	0.31	0.27	1	0.25	0.77	0.40	0.50

Table 7.14 Proportions of genera which have the property [-NC]

Table 7.14 shows that the proportions of genera which have [-NC] are greater in 5 out of the 6 areas. Southeast Asia & Oceania is the only one area that does not conform to this. Although the preference for +{N,NUM,NSG,(X)} shows a relatively small difference (0.95 to 1), the generalisation is unlikely to be valid. According to Dryer (2003: 113), as a rule of thumb, if the tendency is quite strong at least in 5 areas, and the exceptional area shows a small difference, it is acceptable that there is a real linguistic preference. Although some areas, namely Africa and Eurasia seem to show a relatively strong tendency, 2 areas do not clearly show a strong tendency, namely Australia-New Guinea (0.69 to 0.25) and North America (0.90 to 0.77). Since, only 2 areas do not clearly show a strong tendency, the second generalisation “if -{N,NUM,NSG,(X)} then [-NC]” is therefore true but only in a weak sense (cf. Table 7.12).

In sum, between the two hypothesised implicational universals, only the generalisation “if -{N,NUM,NSG,(X)} then [-NC]” is true but in a weak sense. As for the generalisation “if +{N,NUM,NSG,(X)} then [+NC]”, typologically the generalisation is not true, since 2 areas, namely Southeast Asia & Oceania and North America do not conform to the hypothesis. According to this result, this means that a language where {N,NUM,NSG,(X)} is not present is likely to lack the category of noun class. In any case, since there is certain relationship between {N,NUM,NSG,(X)} and noun class in some geographical areas, the next section provides a possible account of how the two features are associated.

7.4 Why is {N,NUM,NSG,(X)} associated with noun class?

The association between {N,NUM,NSG,(X)} and noun class systems seems rather unclear at this stage. Nevertheless, one possible account may be linked to two related issues already mentioned in §5.1.3, namely the use of third person plural pronouns as a source for plural markers and the hypothesis about the development of obligatoriness of plural markers.

The hypothesised scenario is as follows. At first, we may apply Givón's (1976) account for the rise of agreement to the rise of gender. According to Givón, a third person plural pronoun is used as an *anaphoric pronoun* (i.e. the pronoun used to refer back to the subject noun phrase in the sentence), for example, English *the man, he came* (Givón 1976: 155), compared to *the men, they came*. In this case, the pronouns *he* and *they* function as anaphoric pronouns referring back to *the man* and *the men* respectively. The English example shows us that the use of anaphoric pronouns is attributed to the topicalization of the subject noun phrase. Also, from the English examples, the anaphoric pronouns agree with the subject noun phrase in number (*he* versus *they*), and in fact in gender as well, for example, *the woman, she came*. Therefore, at this stage, we can see how anaphoric pronouns are associated with gender and number. That is to say, anaphoric pronouns originated in personal pronouns which may correspond to the gender and number of the noun subjects. However, the English examples do not provide us with concrete evidence for the change from anaphoric pronouns into plural markers or number markers in general. To see how anaphoric pronouns can be reanalysed as number markers, consider the example from Baka, a Niger-Congo language (already illustrated in example (4.4)) once again.

(7.1) *Baka* (Christa Kilian-Hatz, personal communication, in Heine and Kuteva 2002a: 237)

- (a) wósé wó á gɔ
 woman 3PL ASP go
 'The women are going.'
- (b) wósé-o (wó) á gɔ
 woman-3PL (3PL) ASP go
 'The women are going.'

We can now see that via the process of grammaticalization, the anaphoric pronoun *wó* is reanalysed as a plural marker *-ó*. *Baka* provides the evidence of the development of the plural marker from anaphoric pronouns. In addition, it suggests that an anaphoric pronoun can become a plural marker in a language showing characteristics of being highly grammaticalized.¹⁹ Still, this language does not provide evidence which would show how a number marker could become associated with a noun class or gender marker. The evidence supporting the relationship between the three grammatical features can be found in //Ani (Khoisan; Botswana). Consider the examples below.

(7.2) //Ani (Vossen 1986 cited in Heine 1999)

- a. lúí-m' !'úí-mà (-'á) tí mún-m-tà
 one-3M.SG leopard-3M.SG OBJ 1SG see-3SG-PST
 'I have seen one leopard (earlier today).' (p.38)
- b. tí mún-à-tà |ámì kún xèù-tsa 'à
 1SG see-II-PST two big hippo-M.DU OBJ
 'I saw two big hippos.' (p.34)(II=verbal juncture II)

//Ani is a language with three genders, namely masculine, feminine and common (Heine 1999). In this language, it can be observed that the overt gender markers *mà* 'third person-masculine-singular' and *tsa* 'third person-masculine-dual' encode three

¹⁹ This point is also important because both number and gender are categories found in highly grammaticalized languages and non-isolating languages. Also, the fact that it is the category of number which is always found in languages with gender but not vice versa suggests that the languages with gender are expected to be the most highly grammaticalized.

grammatical categories, namely *person-gender-number* (or PGN), all in one morpheme. This reflects the possibility that following the grammaticalization process described in English and Baka, the overt PGN marker might have developed from the third person pronouns. In this language, the third person pronouns are made distinctive in gender and number (Siewierska 2005). //Ani illustrates the final process where the anaphoric pronouns have changed into bound morphemes encoding PGN on nouns. Via the process of lexicalization the bound morpheme may ultimately become an obligatory part of the word (cf. §3.2.2.2). For example, in Kikuyu (Niger-Congo; Kenya; Mugane 1997), the root *-ti* means 'tree'. The root cannot be used alone, rather it must be attached to the class/number prefixes *mu* (Class III, SG) or *mi* (Class IV, PL), yielding *mu-ti* '(a single) tree' or *mi-ti* 'trees'.

Turning to the hypothesis made in §5.1.3 about the development of obligatoriness in grammatical numbers, lexicalized forms such as *mu-ti* 'tree', which denotes gender and number, might have existed before the use of numerals proper in this language (cf. §5.1.3, 6th degree). Once the numerals came into use, then {N,NUM,NSG,(X)} arose in the language.

However, this account is just one possibility and may apply to African languages only. In other groups of languages, such as Indo-European languages and those of Australia-New Guinea and Americas, the question of why non-singular markers are required in CNNCs remains obscure.

7.5 Conclusion

This chapter has moved on to another step of typological tasks, namely *typological generalisation*. The systematic patterns of the distribution of the language types of $CNNC_{NSG}$ which were established in Chapter 6 have now been generalized in accordance with noun class systems. Consequently, an implicational universal has been proposed—that is,

“a language in which {N,NUM,NSG,(X)} is not employed at all tends to lack the category of noun class (even if the language has the category of number.)”

This implicational universal has been tested statistically within a method for testing the statistical significance of implicational universals proposed by Dryer (2003) and it was found that there is a relationship between the absence of noun class and the absence of {N,NUM,NSG,(X)} in the language.

8 Historical Origins of CNNC_{SG}

In Chapter 1, the general assumption that human language basically evolves towards the balance of economy and distinctness was made. Accordingly, to express the number of things economically, humans would be expected to use quantificational expressions consisting of just two linguistic elements, namely a noun denoting things and a cardinal numeral denoting numbers. Hence, the pattern of {N,NUM} should best suit these principles. However, as can be seen from Chapter 6, in a large number of languages, the CNNCs are more complex due to the quantificational extra elements. The complex structures of CNNCs violate any supposed principle of economy in language evolution, since the extra elements do not seem to carry any extra information. Also, due to the fact that the number of quantified referents is already identified by the numerals, the principle of distinctness does not seem to account for the complexity. So, how and why did the other patterns of CNNCs, especially those of complex structures, arise in a number of human languages?

The typological generalization made in Chapter 7 helps us to understand only that the major language types of CNNC_{NSG}, namely {N,NUM}, {N,NUM,NSG,(X)}, and {N,NUM,CLF}, tend to have a correlation with classifier systems. In other words, the variety of structures is no more or less attributable to the existence of classifier systems. However, the correlation between CNNCs and classifier systems does not answer the key question raised above. The *typological perspectives* alone then may not be sufficient to provide insights into the diversity of CNNCs, and particularly the structural complexity of CNNCs.

We will therefore turn to *diachronic perspectives* to trace the development of the patterns of CNNCs with the ultimate aim being to understand how these patterns come into use in human languages. Therefore, the thesis will investigate historical developments of the patterns of CNNCs and the motivations for these developments by using *evidence from old texts* as well as *theoretical (diachronic) approaches*. See

Chapter 3 for a discussion of the application of evidence from old texts and a brief review of theoretical (diachronic) approaches.

Representing the historical developments of CNNCs, the symbol ‘<’ meaning ‘derives from’ is used. For example, the notation ‘{X} < {Y}’ indicates that the structural pattern {X} derives from the structural pattern {Y}. Note that each diachronic pattern represents one of the possibilities only; there are probably other possibilities as well. That is to say, for the notation ‘{X} < {Y}’; {X} may also derive from other patterns in addition to {Y}; and {Y} may turn into other patterns besides {X}. For example, in contemporary Mandarin Chinese, there exists the pattern of {N,NUM}. This construction has two origins: one has developed from {N,NUM,CLF} and the other is the original structure inherited from Classical Chinese (Tao 2005: 287).

Because the descriptions of the historical origins of $CNNC_{SG}$ and $CNNC_{NSG}$ are rather lengthy, this chapter touches only on the historical origins of $CNNC_{SG}$. For the historical origins of $CNNC_{NSG}$, the reader is referred to Chapter 9. In Chapter 8, the historical development of the various structural patterns of current $CNNC_{SG}$ (as illustrated in Chapter 6) is explored. However, only the structural types of $CNNC_{SG}$ for which the evidence is satisfactory are reported. These include {N,NUM} (§8.1), {N,NUM,SG} (§8.2), {N,NUM,CLF} (§8.3), and {N,NUM,CLF,SG} (§8.4).

8.1 {N,NUM}

The pattern of {N,NUM} is the simplest pattern of $CNNC_{SG}$, ignoring the pattern of {N,SG} where the singular or singulative marker is not a cardinal numeral as such. The pattern of {N,NUM} is composed of only two core constituents, namely a noun and a numeral, and in such case one would not suppose it to derive from any other constructions. Despite this, the pattern of {N,NUM} may derive from other constructions, namely {N+ ‘one-ness’} (§8.1.1) and {N,NUM,CLF} (§8.1.2).

8.1.1 {N,NUM} < N + ‘one-ness’

This historical pattern indicates that the pattern of {N,NUM} may develop from a syntactic construction in which the noun was modified by words meaning ‘one-ness’, such as ‘alone’ and the like, through the process of numeralization (cf. §4.4). This historical pattern is supported by diachronic and synchronic evidence. For example, the Indo-European roots for the numeral one, i.e. *oi-* and *sem-* originally mean ‘alone’ and ‘together’ (i.e. unified into a single element) respectively (Burrow 2001: 285). The polysemous words with a numerical interpretation for ‘one’ observed in present-day languages may mean ‘alone’, ‘small’, or ‘little finger’ and the like. Such words may be found in Kwazá (Kwaza; Brazil), where the word *tei* means ‘to be one’ and ‘alone’ (Voort 2004: 214); in Pirahã, where the word *hói* means ‘small size’ (Everett 2005: 623); and in Haruai (Upper Yuat; Papua New Guinea), where the word *agly* means ‘little finger’ (Comrie 1999: 91)). The topic of words with a numerical interpretation has already been discussed in Section 4.4, so it will not be repeated here. The stage of *N+* ‘one-ness’ can be regarded as an embryonic stage prior to the emergence of {N,NUM}.

This change is perhaps attributed to the user’s need to express the exact number (i.e. *one*). That is, the change involves the principle of expressivity. However, the principle of economy also plays a role in this matter. Instead of creating a new linguistic symbol, humans may prefer to make use of an internal source to refer to a similar concept. Metaphorical extension such as this is a usual phenomenon giving rise to the polysemy of words. In sum, the overall change deals with semantico-syntactic change

(i.e. a change involving both semantic and syntactic aspects), as the semantic change makes the syntactic construction like *N+ 'one-ness'* turn to {N,NUM}.

8.1.2 {N,NUM} < {N,NUM,CLF}

This pattern indicates that {N,NUM} or (to be more precise) {N,NUM.CLF} may develop from {N,NUM,CLF}. {N,NUM.CLF} is a simplex construction where the concepts and forms of a numeral and a numeral classifier are fused into one morpheme and are mostly morphologically unanalyzable, indicated by the notation 'NUM.CLF'. *Formally*, the pattern of {N,NUM.CLF} may be interpreted or reanalysed as {N,NUM} because the classifier is fused with and seemingly lost in the numeral. In other words, the users of the language may treat the fused form of *NUM.CLF* as a sheer numeral, presupposing that the language has more than one set of numerals for counting different classes of things. Yurok (Algic; California) is an example of such a language. From the examples below, it can be noticed that it is very difficult (if not impossible) to separate the numeral 'one' from the numeral classifiers.

(8.1) *Yurok (Blevins 2004: 1)*

(a) *k.jh.t.ɿʔy.ʔy* *pu:k*
one.CLF *deer*
 'one deer'

(b) *kohtoh* *haʔa:g*
one.CLF *rock*
 'one rock'

In some languages such as Yurok, the process of fusion is fully complete. In other languages, nonetheless, the process is still on the way, and hence provides the helpful evidence of change in progress. Beijing Mandarin Chinese is an example of such a language.

According to Tao (2005), Beijing Mandarin currently has 3 structural types of CNNC_{SG}, namely {N,NUM,CLF}, {N,NUM}, and {N,CLF}. In this section, only

{N,NUM} concerns us. The pattern of {N,NUM} is exceptional and rare in Chinese. There are two origins of {N,NUM}. One is inherited from ancient Chinese (henceforth *the inherited {N,NUM}*), and the other is newly developed from {N,NUM,CLF} (henceforth *the new {N,NUM}*). The inherited {N,NUM} is usually used in formal writing, and the latter in colloquial speech. Another difference is that the numeral *yi55*²⁰ ‘one’ in the inherited {N,NUM} follows the *tone sandhi* rules (rules about a change in lexical tones when different tones are close together), whereas the numeral ‘one’ in the new {N,NUM} does not. As the new {N,NUM} represents a change from {N,NUM,CLF} to {N,NUM}, it will therefore be focused on here.

The development of the new {N,NUM} from {N,NUM,CLF} can be explained by a phonological change occurring in the everyday use of language. The change takes place when the numeral ‘one’ is followed by the classifier *ge51*, a general classifier for most nouns including human nouns, roughly translated as ‘item’. Tao (2005: 308) refers to the phonological change leading to the syntactic effects on the whole noun phrase as “phono-syntactic conspiracy”, and describes the phonological change as follows. The notations in the square brackets summarize the process. Note that the symbol (>) represents ‘becomes’.

(1) Following a tone sandhi rule, the tone of the numeral *yi55* ‘one’ changes to *yi35* when preceding the numeral classifier *ge51*. [*yi55* > *yi35*]

(2) Because *yi35* and *ge51* usually come together, the two elements form one phonological unit (i.e. virtually one word). [*yi35 ge51* > *yi35ge51*]

(3) Then the word stress (´) falls on the numeral *yi35* ‘one’; and the numeral classifier *ge51* adopts the neutral tone, becoming *ge*. [*yi35ge51* > *yi35ǵe*]

(4) Being unstressed, the vowel *lel* in *ge* is reduced to schwa *lǝ*. [*yi35ǵe* > *yi35ǵǝ*]

²⁰ The numerals appearing together with the word represent the lexical tone of the word, e.g. (*yi*)*55* refers to the high-level tone (of the word *yi*).

(5) Then the intervocalic consonant (a consonant appearing between the vowel sounds, i.e. /g/ here in *yi35ge*) is lost. The two syllables are reanalysed as one chunk. So now *yi35ge* becomes *yi35ə*. [*yi35gə* > *yi35ə*]

(6) The vowel sequence (i.e. /ə/ here) is deleted and finally the classifier *ge51* is completely dropped, leaving the numeral *yi35* (formerly *yi55* high-level tone) with a high-rising tone. [*yi35ə* > *yi35*]

The tone *yi35* is a frozen tone—that is, the tone does not follow the Mandarin tone sandhi rules because it implicitly has the word *ge51*. Although the classifier *ge51* is formally eliminated, the numeral classifier is retained semantically. This is the reason why the numeral *yi35* ‘one’ does not require the numeral classifier *ge51*. Therefore, the grammatical status of the numeral *yi35* ‘one’ lies between an *allomorph* (i.e. an alternating form of a morpheme) of the morpheme *yi55* and the new morpheme *yi35* denoting ‘one item of’. Tao provides two examples of minimal pairs illustrating the grammatical meaning of the tone (*yi*) 35:

(8.2) *Beijing Mandarin* (Tao 2005: 310)

(a) *yi35 ber214*
one notebook
‘one notebook’

(b) *yi51 ber214*
one notebook
‘one copy (of a book, a notebook)’.

The *ber214* ‘notebook’ in (8.2b) is treated as a numeral classifier for books and notebooks because there is no numeral classifier appearing in the phrase, whereas the *ber214* ‘notebook’ in (8.2a) is treated as a noun due to the fact that *yi35* implicitly denotes the numeral classifier *ge51* ‘item’, and so the noun *ber214* is not treated as a numeral classifier.

Chirkova (2004: 1) points out that *yi35* is in the process of *lexicalization* (cf. §3.2.2.2). The concepts of the numeral *yi55* and classifier *ge51* are fused into one morpheme, giving rise to the new form *yi35* with the new meaning ‘one item of’—that is, the new word in the lexicon. The frequency of use of the new numeral *yi35* is increasing. Yet, it is not completely lexicalized, since it shares the character transcription with *yi55* ‘one’—in other words, it has not yet been assigned its own character transcription and is hence not included in dictionaries. However, due to the fact that the construction *yi55+ge51* is reanalysed as the numeral assuming the grammatical function of a numeral classifier, the process can then also be regarded as *grammaticalization* in some places (see, for example, Tao 2005; Wischer 2006: 129). The Chinese *yi35* no longer requires the numeral classifier *ge51*, since the frozen tone has assumed this function already. In sum, the fusion of numeral and classifier is the product of the intersection of lexicalization and grammaticalization.

Regarding the motivations for the change, the reason why the numeral classifier *ge* has been dropped is not because of its redundancy, rather because the speaker presumably wants to minimize the effort in articulation. Therefore, the underlying motivation for the change involves the principle of economy in language. This can happen when the words in question are highly frequently used. Tao (2005: 309-310) attributes this phono-syntactic change to the relatively high frequencies of *yi55* and *ge51* in everyday conversation.

According to the current sample, the simplex construction {N,NUM.CLF} is not unusual in human languages. In addition to Yurok and Beijing Mandarin, there are several instances observed, such as Nivkh (isolate; Siberia, Russia; Gruzdeva 1998: 24), Warekena (Arawakan; Brazil and neighbours; Aikhenvald 1998), and Teribe (Chibchan; Costa Rica and Panama; Quesada 2000). To illustrate, in Teribe, there are 6 classes of numerals varying according to the classes of nouns. However, the classifiers are not yet fused with but only prefixed to the numerals as shown below.

	animate	round objects	wide objects	long objects	long-wide objects
'one'	kl-ara	kw-ara	k-ara	pl-ara	kri-na

Table 8.1 Six classes of numeral 'one' in Teribe

It should be remarked that some languages have a separate set of numerals for counting humans. Nevertheless, these languages cannot be treated in the same way as Yurok or Beijing Mandarin. Good examples are provided by Irish and Tamil (Dravidian; Southern India). Both have only one class of nouns treated specially, i.e. humans, and therefore they are not regarded as numeral classifier languages. The numerals for counting humans in both languages already existed in their parent languages, e.g. Classical Tamil (150 B.C.- 6th A.D.) *oruvar* 'one.person', *iruvar* 'two.person' (Rajam 1992: 291); Old Irish *óenar* 'one person', *dias* 'two.person, couple' (McCone 2005: 62). In both earlier languages, no numeral classifiers are mentioned in the old texts (cf. Rajam 1992 and McCone 2005), so the special numerals should not be regarded as the remnants of ancient numeral classifiers. The examples below show that there are two numeral forms for non-humans and humans respectively.

(8.3) *Modern Irish (Doyle 2001)*

(a) *trí bhád*
three boat
'three boats' (p.55)

(b) *triúr saighdiúir*
three.person soldier
'three soldiers' (p. 56)

(8.4) *Modern Tamil (Lehmann 1989)*

(a) *oru praricu*
one gift
'one gift' (p. 115)

(b) *paiyan oruvan*
boy one.male.person
'one boy' (p.114)

(c) *oruvan* *va-nt-aan*
one.male.person *come-PST-3SG.M*
'one male person came.' (p. 114)

The special numeral forms in Irish and Tamil are employed to refer to people, just like a pronoun as in (8.4c). In Tamil, the cardinal numerals of this kind are thus referred to as “pronominalized cardinal numerals” (cf. Lehmann 1989: 113) or “numeral pronouns” (Rajam 1992: 290); that is, they behave nominally as a pronoun. In Irish they are said to be able to stand alone without accompanying nouns (Doyle 2001: 56).

8.2 {N,NUM,SG}

{N,NUM,SG} is a complex construction. It is conjectured to derive from {N,SG} which in turn develops from two different historical sources. The first source is {N,NUM} and the second source is a noun combined with a third person singular pronoun.

8.2.1 {N,NUM,SG} < {N,SG}<{N,NUM}

This pattern indicates that {N,NUM,SG} may develop from {N,SG} which in turn derives from {N,NUM}. This historical path is illustrated by Burushaski (isolate; Pakistan) and Pame (Oto-Manguean; Mexico). We will consider Burushaski first, since it provides a clear example. Then we will move to the less clear case of Pame.

8.2.1.1 Burushaski

There are three structural patterns of CNNC_{SG} in Burushaski, namely {N,NUM}, {N,SG}, and {N,NUM,SG}, as exemplified in (8.5 a-c) below.

(8.5) *Burushaski (Lorimer 1935)*

a. {N,NUM}

hik²¹ bΔda

one pace, step

'one pace, step' (p.191)

b. {N,SG}

sisΔn

person (PL)-SG

'(a) person' (p.48)

c. {N,NUM,SG}

hΔn huyεsΔn

one goat, sheep (PL).SG

'a goat, sheep' (p.48)

²¹ Note that the numeral *hik* 'one' has another two variants, namely *hΔn* and *hin*.

Since there is no historical evidence, we need to use a theoretical diachronic approach. Considering the variety of structural patterns of $CNNC_{SG}$ in the language, it seems Internal Reconstruction (cf. §3.2.2.3) would suit a case like Burushaski. In this case, Grammaticalization (cf. §3.2.2.1) will be used as a technique to reconstruct the developmental pathway of the current $\{N,NUM,SG\}$. Within the grammaticalization framework, it is conjectured that the three structural patterns of $CNNC_{SG}$ used in contemporary Burushaski have derived from the same original pattern, namely $\{N,NUM\}$. These patterns are believed to reflect various stages of change. The process is shown as in Figure 8.1.

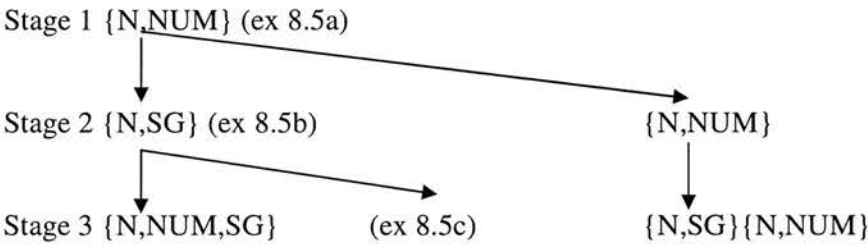


Fig. 8.1 Process of change in $CNNC_{SG}$ in Burushaski
NB: ex = example

Stage 1

According to the grammaticalization principle, the earlier the linguistic forms are, the more lexical they are expected to be. Based on this principle, the $CNNC_{SG}$ in (early) Burushaski or Pre-Burushaski is conjectured to be composed of a noun and the numeral *one* (i.e. $\{N,NUM\}$) as in (8.5a) above. In other words, the example in (8.5a) represents the structural pattern inherited from the early stage of the Burushaski $CNNC_{SG}$.

Stage 2

$\{N,NUM\}$ has split into two types, namely $\{N,NUM\}$ and $\{N,SG\}$. The pattern of $\{N,SG\}$ (cf. the example 8.5b) which is a marginal case of $CNNC_{SG}$ is regarded as an intermediate stage between $\{N,NUM\}$ and $\{N,NUM,SG\}$. $\{N,SG\}$ has emerged because

the language has nouns which are semantically plural or in some cases are referred to as *plural nouns* or *collective nouns*. When these nouns are required to be counted individually, the numeral ‘one’ is then used in a grammaticalized form as a singulative marker. The development of the singulative marker from the numeral ‘one’ is evidenced by the similarity between meaning and form. In this language, the singulative suffixes *-ʌn* or *-ɛn* seem to be associated with the numeral *hʌn* ‘one’ (Lorimer 1935: 46) (cf. §5.1.2).

Stage 3

After the nouns which are semantically plural (such as *sis*, ‘people’) have gained a singular form, the speaker uses the numeral ‘one’ plus the singular form when they are enumerated as in (8.5 c), while the two preceding structural patterns also remain in use.

In sum, the structural pattern of {N,NUM,SG} has come into being because some nouns are semantically plural. The mechanism involved in this development is grammaticalization.

8.2.1.2 Pame

According to Manrique Castanade (1967), contemporary Pame has 3 structural types of CNNC_{SG}, namely {N,NUM}, {N,SG}, and {N,NUM,SG} as illustrated in (8.6 a-d).

(8.6) *Pame (Manrique Castanade 1967)*

(a) {N,NUM}
nada kuhú
one sorcerer
 ‘one sorcerer’ (p. 345)

(b) {N,SG}
na-çi°
SG-tooth (PL)
 ‘tooth’ (p.346) (NB: *na-* = *pro-clitic*)

- (c) $\{N,SG\}$
n-çi^o
SG-tooth(PL)
'tooth' (p.346) (NB: n- = (dead) prefix (i.e. no longer used as a prefix))
- (d) $\{N,NUM,SG\}$
nada n-tao
one SG-eye
'one eye' (p.346)

Like the case of Burushaski, due to the lack of historical evidence, we need to use IR with Grammaticalization as a technique to reconstruct the historical development of the current pattern of $\{N,NUM,SG\}$. Consistent with the grammaticalization framework is the supposition that the three structural patterns of $CNNC_{SG}$ used in contemporary Pame have derived from the same original pattern, namely $\{N,NUM\}$. These patterns are believed to reflect various stages of change. The process is shown as in Figure 8.2.

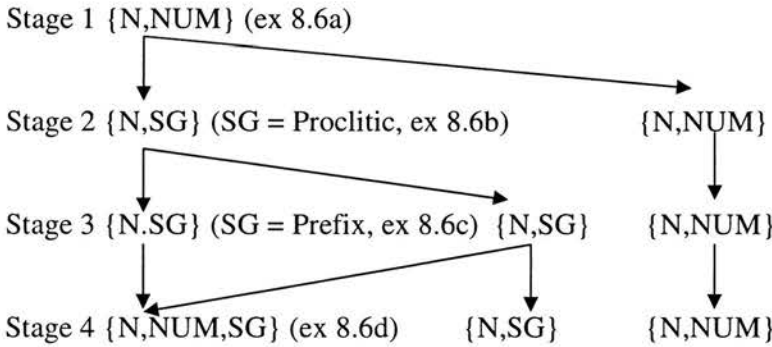


Fig. 8.2 Process of change in $CNNC_{SG}$ in Pame

Stage 1

Based on the grammaticalization principle that the earlier linguistic forms are likely to be more lexical (or less grammatical), the CNNC_{SG} in (early) Pame or Pre-Pame is conjectured to be composed of a noun and the numeral *nada* ‘one’ (i.e. {N,NUM}) as in (8.65a) above. In other words, the example in (8.6a) represents the structural pattern inherited from the early stage of the Pame CNNC_{SG}.

Stage 2

{N,NUM} might have remained in use. The proclitic *na-* came into use to mark singularity of nouns, for example, *na-šthe* ‘SG-nopal’. It is possible that the proclitic *na-* was grammaticalized from the numeral *nada* ‘one’. This is because the form and meaning of the proclitic *na-* and *nada* seem to be related. Also, grammaticalization of the singular marker from the numeral ‘one’ is fairly common. Burushaki provides a clear example. The reader is referred to Heine and Kuteva (2002a: 223-224) for more examples of the development of singular or singulative markers from the numeral *one*. The structure is encoded as {N,SG}. Note again that this type is regarded as a marginal case of CNNC_{SG}, since it does not contain the numeral *one* proper. In any case, {N,SG} can be regarded as an intermediate stage between {N,NUM} and {N,NUM,SG}.

Stage 3

{N,NUM} and {N,SG} (where SG is a proclitic) might have remained in use. At this stage, the prefix *n-* has appeared in order to denote the singularity of nouns. According to grammaticalization theory, an affix is often argued to develop from a clitic (e.g. Givon 2001: 55). Considering the similarity of form and meaning, it is possible to conjecture that the prefix *n-* might have developed from the proclitic *na-*.

Stage 4

{N,NUM} and {N,SG} (where SG is a proclitic) might still have remained in use. The prefix *n-* has been fused with the root and is no longer a prefix, e.g. *n-çi?* ‘SG-tooth’ > *nçi?* ‘tooth’; *n-tao* ‘SG-eye’ > *ntao* ‘eye’. The new words *nçi?* ‘tooth’ and *ntao* ‘eye’ illustrate the dead prefix *n-* which has been fused into the nouns. The prefix+noun

construction has been fully lexicalized, as evidenced by the fact that the lexicalized form requires the proclitic *na-* or the numeral *nada* ‘one’ when the singularity is emphasized, e.g. *na-nçi* ° ‘SG-tooth’; *nada ntao* ‘one eye’. The process whereby the prefix has been fused with the noun corresponds to *lexicalization* (cf. §3.2.2.2). In fact, when the prefix is reanalysed as a part of the word, the structural pattern such as *nada ntao* ‘one eye’ should be regarded as {N,NUM}. However, it is more useful here to label this construction as {N,NUM,SG}, since it may help understand how the unmotivated construction (in other languages) came about.

In relation to the motivation for the change, it is not clear from the source what the motivation is for the emergence of {N,SG} as an alternative form. However, the reason might not be different from the case of Burushaski. That is, via the grammaticalization process, the numeral *one* also became used as a singulative marker for the collective nouns. We may need to check whether the nouns attached by the prefix or proclitic are nouns that are often found in the plural form, such as those meaning *eyes* or *teeth*. If this is the case, the motivation clearly involves quantification.

In sum, with the current structural patterns of CNNC_{SG}, it is conjectured that the Pame {N,NUM,SG} has developed from {N,SG} which in turn has developed from {N,NUM} via the grammaticalization process.

8.2.2 {N,NUM,SG} < {N,SG} < N+third person singular pronoun

This pattern indicates that {N,NUM,SG} may develop from {N,SG}. An example is taken from Lunda (Niger-Congo; Democratic Republic of Congo). In Lunda (Kawasha 2003), the CNNC_{SG} is {N,NUM,SG} as shown in (8.7).

- (8.7) *Lunda* (Kawasha 2003: 124)
- | | |
|---------------------|--------------|
| <i>mu-ntu</i> | <i>wu-mú</i> |
| <i>I-person</i> | <i>I-one</i> |
| <i>‘one person’</i> | |

Within the grammaticalization framework, it is hypothesized that the structural pattern of {N,NUM,SG} has developed from {N,SG} which in turn has developed from *N+classifier*. The hypothesis is also based on the tendency examined in Burushaski (cf. §8.2.1.1) and Pame (cf. 8.2.1.2). The process of change is depicted as in Figure 8.3 below.

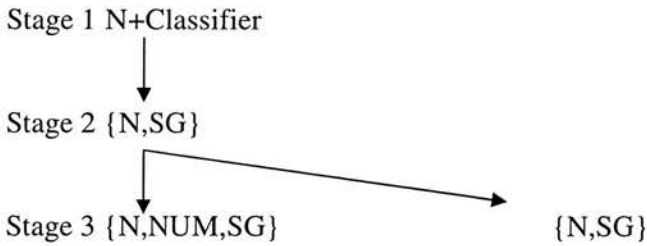


Fig. 8.3 Process of change in CNNC_{SG} in Lunda

Stage 1

It is speculated that the singular marker might have been required because nouns were vague in number and because the numeral ‘one’ was not available. Considering the case of plural number in African languages, it is possible that the singular marker might have been derived from the third person pronoun as well.

In many African languages, 3rd person plural pronouns have been added to nouns and have developed into nominal plural markers [...] In accordance with this strategy, the personal pronoun follows the specified unit. (Heine and Reh 1984: 234)

Niger-Congo semantic classification involves multilateral oppositions: humans, animals, plants, paired body parts, mass nouns and liquids, abstracts, and others [...] The affixes which mark these oppositions appear throughout Niger-Congo; apart from those for singular and plural person classes, which often resemble the third person pronouns, they have no discernible etymology (Williamson and Blench 2000: 13)

If this is the case, it is speculated that the fusion of noun and singular marker might have happened before the use of the numeral ‘one’. For the process of the

development of noun class (expressed along with number) from the third person pronoun, refer to §7.4. Even though Section 7.4 concentrates on the plural number, the singular number might undergo the same process. That is, the grammaticalization might begin with a noun combined with an anaphoric pronoun, and then the anaphoric pronoun might grammaticalize into the number/gender marker. Finally, the marker would be fused

Stage 2

Like Pame, the noun class/singular marker is fused with the noun as an obligatory part of the noun via the process of lexicalization. So, the root cannot be used independently but must be bound with a noun class prefix via this process. For example, in *mu-ntu* (I-person) 'person', the speakers cannot say **ntu* only. However, we do not know exactly where the *mu-* is derived from.

Stage 3

Once the numeral *one* was introduced into the language, the numeral was combined with the nouns which have the singular marker obligatorily bound to the noun.

To sum up, it is conjectured that before the emergence of CNNC_{SG} in Lunda (and perhaps Niger-Congo languages in general), or at the time when numerals had not been introduced into the language, there might have been a basic device to distinguish singular and plural number of referents in pre-Lunda at least. At the first stage whereby the numerals were combined with nouns to form CNNC_{SG} in the language, it is more likely that the nouns might have already been affixed with the noun class/number. In other words, it should not be the case that the nouns were redundantly affixed by the noun class/number after the CNNC_{SG} had been established.

8.3 {N,NUM,CLF}

{N,NUM,CLF} may develop from the patterns of {N,NUM} (§8.3.1) and {N,NUM,SG} (§8.3.2). Note that the development of {N,NUM,CLF} to be demonstrated applies to both $CNNC_{SG}$ and $CNNC_{NSG}$.

8.3.1 {N,NUM,CLF} < {N,NUM}

This pattern indicates that {N,NUM,CLF} may develop from {N,NUM}. This developmental pattern is illustrated by Chinese (the cover term we use here to refer to various Sinitic languages), for example, Mandarin Chinese, Cantonese), Japanese (isolate) and Khmer (Austro-Asiatic; Cambodia). Among classifier languages, Chinese is the only language that illustrates the complete origin and development of a numeral classifier system. Since it provides comprehensive data, the Chinese case deserves a more detailed discussion. Japanese and Khmer are numeral classifier languages where the pattern presumably developed due to language contact.

8.3.1.1 Chinese

The typical structural pattern of CNNCs in Chinese is {N,NUM,CLF}. According to the evidence from the ancient written records of the Archaic period (14th - 2nd BC), the pattern of {N,NUM} is the basic pattern of CNNCs as shown in (8.8a-d) (8.8a and 8.8c for $CNNC_{SG}$; 8.8b and 8.8d for $CNNC_{NSG}$). The order of numerals with respect to the noun they modify is flexible. In 8.8a-b the numeral precedes the noun, whereas in 8.8c-d the noun precedes the numeral. However, the latter is the typical order (Peyraube 2004: 988; Wang 1994: 90; Dobson 1962: 26).

- (8.8) *Chinese (Archaic)*
- (a) *wu* *he* *ai* *yi* *niu*
 1SG *why* *begrudge* *one* *ox*
 ‘*why (should) I begrudge one ox?*’ (Peyraube 2004: 1001)

(b) *nai bu san gui*
 then divine three turtle
 ‘Then (used) three turtles to divine.’ (Wang 1994: 90)

(c) *shing niou i*
 red ox one
 ‘one red ox’ (Dobson 1962: 26)

(d) *niou ell*
 ox two
 ‘two oxen’ (Dobson 1962: 26)

There exists, however, a syntactic construction which is similar to {N,NUM,CLF}. This construction is composed of a noun, a numeral, and a measure word (MW). The structural pattern can be represented as N+Num+MW, where the noun precedes the numeral and the numeral precedes the measure word as shown in (8.8 e-f) below.

(e) *ma er shi cheng*
 horse two ten MW (for four horses)
 80 horses (Wang 1996: 84)

(f) *jeu-chanq ell yeou*
 sacrificial wine two flask
 ‘two flasks of sacrificial wine’ (Dobson 1962: 29)

The measure words as shown in (8.8e) and (8.8f) are referred to by some Chinese grammarians as *liangci* ‘measure word’ (Wang 1994: 78). The function of *liangci* is to unitize a group of referents for counting as in (8.8e) and to unitize a mass referent for measuring as in (8.8f). In (8.8e), the horses are grouped in four (i.e. *cheng*). What are counted are *groups* of horses, not *the horses*. In (8.8f), *wine* is put into a unit of measure with a flask so that it will be countable. So, what are counted are *flasks*, not *wine*.

There are also a few measure words used with particular *individual* nouns, for example, *pi* for horses, *leang* for chariots, *ren* for humans, *bor* for humans of higher rank, and *ge* for various items (including humans) (Dobson 1962: 28; Wang 1996: 82). Examples (8.8g)-(8.8h) are below.

(g) *maa* *bae* *syh* *pi*
 horse *hundred* *four* *MW*
 ‘one hundred and four horses’ (Dobson 1962: 28)

(h) *che* *san* *shyr* *leang*
 chariot *three* *ten* *MW*
 ‘thirty chariots of war’ (Dobson 1962: 28)

This type of measure words are referred to as “individual measure” in Chao (1968: 584). Due to the fact that the individual measures do not seem to classify nouns in terms of inherent properties, such as animacy or shape, and some individual measures are for nouns with different semantic categories, scholars do not regard them as numeral classifiers. Wang (1996: 68), however, considers these individual measures as “plausible classifiers” and the construction containing the individual measures as an embryonic stage for {N,NUM,CLF} in the later periods.

Moreover, there exists the structure which makes use of the repetition of the noun being counted after the numerals as illustrated by the examples (8.8i-j). This structure may be referred to as a “repeater construction” and the noun repeated in this construction is called a “repeater” or “echo-classifier” (Aikhenvald 2000: 103)²². The words *ren* and *niou* in (8.8i-j) are examples of a repeater. There are other nouns of this kind, such as *tian* ‘field’, *yang* ‘sheep’, *ren* ‘person’, *niou* ‘ox’, *tuo* ‘bamboo’, and *gwo* ‘scalp’. This pattern can be observed even in the early Archaic period (Dobson 1962: 28-9; Wang 1996: 71) as shown in the examples (8.8 i-j).

²² Regarding the use of echo-classifiers, Jones (1970: 2) points out that the repeaters are required at the early stage of numeral classifier systems due to “an inadequate supply of classifiers.”

(i) *niou san bae wuu shyr wuu niou*
ox three hundred five ten five ox
 'three hundred and fifty five oxen' (Dobson 1962: 28)²³

(j) *fu ren shi you liu ren*
catch person ten and six person
 'caught sixteen people' (Wang 1996: 69)

Scholars agree that although some patterns are similar to the pattern of {N,NUM,CLF}, the true numeral classifiers which are obligatorily used in the pattern of {N,NUM,CLF} had not yet developed until around the Han period (1st century BC) (Norman 1988: 115; Wang 1996: 78; Peyraube 2004).

In the Han period, the (potential) numeral classifiers are more widely observed and show a remarkable increase from 15 in the late Archaic period to approximately 50 in the Han. Since then, the pattern of {N,NUM,CLF} has evidently developed. The use of numeral classifiers became more systematic and consistent; and the widespread use of the pattern of {N,NUM,CLF} gradually replaced the pattern of {N,NUM}. The development of numeral classifier system was completed around the 6th century AD (Wang 1996: 85; Peyraube 2004: 1010). Because the general tendency of Chinese is to put the modifier after the nouns modified, the classifier has moved into prenominal position (i.e. before quantified nouns): N+Num+CLF > Num+CLF+N. This also applies to the construction with the measure words N+Num+MW > Num+MW+N (Peyraube 2004: 1010). In present-day Chinese, the pattern of {N,NUM,CLF} is regarded as a standard pattern of CNNCs, but the pattern of {N,NUM} can also be observed in certain contexts as in (8.9) (cf. §8.1.2).

²³ One may question whether or not the construction containing repeaters such as in (i-j) is applied only to the complex numerals, like 'ox three-hundred plus five-ten-five ox'; that is, where there are two numeral phrases combined. If this is the case, the 'ox' should not be interpreted as a repeater. However, an example from Jinwen bronze script, *tian qi tian, ren wu ren* [field-seven-field, person-five-person] 'seven fields, five people' (Wang 1996: 71), confirms that the repeaters exist.

(8.9) *Chinese (Modern)*

yi huo-ji tui men jin lai
one employee push door enter come

'An employee opened the door and came in'. (Tao 2005: 16)

From the history of Chinese {N,NUM,CLF} just described, the immediate question that arises is: 'How and why did Chinese develop the complex pattern of CNNCs as {N,NUM,CLF} instead of using the simpler one like {N,NUM} which had been already employed?'

Previous studies focus on the chronological development of {N,NUM,CLF}, but pay less attention to the motivations for the change. However, the historical facts about the CNNCs in Chinese as just reviewed above are useful for speculating about the motivations. Based on the history of {N,NUM,CLF}, the potential numeral classifiers (e.g. *pi* for horses) are initially used to individualize some referents which are culturally collectives. These referents are discrete spatial objects but are culturally perceived as a group (e.g. shells and jades) (Wang 1996). Therefore, they are not able to be put directly with numerals. They are rather needed to make things countable or enumerable (cf. §5.2.4). To sum up, the emergence of numeral classifiers in Chinese is attributed to individualization of some collective nouns.

The question remains: how did the use of numeral classifiers spread to individual nouns with a discrete boundary in later periods. It is conjectured here that syntactic analogy plays an important role in making other nouns require numeral classifiers as well. For example, the noun 'ox', which according to the evidence, is regarded as a discrete object, such as 'one ox' below:

(8.10) *Chinese (Archaic)*

shing niou i
red ox one

'one red ox' (Dobson 1962: 26)

However, it is probably because of the structural syntactic analogy of the existing pattern of N+Num+MW, and therefore the pattern of {N,NUM} like 'one ox' turns to N+Num+MW (so like 'ox one ox'; see the example (8.8i)). The question is why the analogy happens. This is perhaps because there are some nouns representing discrete spatial entities (such as *horse*, *shell*, *jade*) which require measure words when enumerated.

The next question is how such a small group of nouns can change the whole system, or in fact the type of language (i.e. from a non-classifier language to a classifier language). In response to this question, the regularization is possible because the word order of Chinese {N,NUM} is flexible, that is, the word order can be either NUM+N or N+NUM (the latter being regarded as a subsidiary pattern, though). The word order of N+NUM when used might have been regularized to have MW by analogy with the existing structure N+NUM+MW, such as *horse-two-MW*. Once the speaker wished to say 'one ox', some would say *one-ox* but others who used *ox-one* would say *ox-one-MW*. In other words, the change takes place in the pattern of N+NUM (not NUM+N). The change may be represented as N+NUM > N+NUM+MW/CLF.

The change happened gradually as evidenced by the variation in the structural patterns of CNNCs in the Archaic period. For some nouns which did not yet have their own classifier, measure words would be used as a numeral classifier. In many cases, the nouns themselves were used to create their own classifier (i.e. repeaters) as can be seen in numeral classifier languages even nowadays, when numeral classifiers are available for new nouns.

Another observation about the change is that the use of numeral classifiers in the early stage is inconsistent and not based on semantic categorization as a numeral classifier system is supposed to be. This is because the nouns in Archaic Chinese do not generally need individualization (i.e. they can be directly accompanied by numerals) and thus the noun classification was not necessary. A numeral classifier was needed just because the pattern of N+NUM+MW forced those nouns to have MW. Therefore, at the initial stage, the nouns could have more than one classifier and vice versa.

(8.12) *Japanese (Modern)* (Kyoko Otsuki, personal communication)

- (a) *hyaku* *wa* *no* *tori* *no* *koe*
hundred *CLF* (lit. 'wing') *ATTR* *bird* *GEN* *voice*
'the voice of hundred birds.'
- (b) *ippon*²⁵ *no* *matsu*
one.CLF *ATTR* *pine*
'one pine'

The origin of the classifier system in Japanese remains problematic. This is because the numerical system has been present in Japanese since the earliest texts. The written records suggest that the nouns are classified on the basis of shape. This implies that the numeral classifier system is likely to be fully developed. So we cannot trace back how the system had been formed and developed prior to its presence in the 8th century. Two speculations have been proposed. One is that the system is borrowed from Chinese. The other is that the system is native to Japanese and became dramatically widespread upon the adoption of Chinese. As for the former speculation, the supporting evidence is that the majority of numeral classifiers used in modern Japanese are often combined with the Sino-Japanese numerals, but are very rare with native Japanese numerals (Downing 1996: 44, 46-47).²⁶ One of the borrowed features from Chinese is numerals which are described as *Sino-Japanese numerals* (e.g. *ichi* 'one'). The Sino-Japanese classifiers which are usually used with the Sino-Japanese numerals form the majority group of numeral classifiers in Japanese (Downing 1996: 35-51). This suggests that the development of classifiers may be attributed to the borrowing of numerals. As for the latter argument, the supporting evidence is the use of a bound form of a numeral classifier for human nouns such as *hito-ri* 'one person'. Considering the use of a numeral classifier-like form in combination with numerals as pronominalized numerals (as in Tamil and Irish), we may see that Japanese *hito-ri* probably represents the same situation.

²⁵ The *Ippon* is derived from *ichi* 'one' + *hon* 'classifier for long and slender objects' through the assimilation process (Downing 1996: 48).

²⁶ Note that Japanese has two systems of numerals, namely native numerals and Sino-Japanese numerals.

Unfortunately, Japanese is a language isolate; its sister languages are dead, so we cannot use the comparative method for reconstruction in this case. Although due to insufficient evidence we may not be able to claim that {N,NUM,CLF} in Japanese changes from {N,NUM}, the presence of the CNNCs without numeral classifiers as shown earlier leads us to think that the structural change from {N,NUM} to {N,NUM,CLF} is possible.

8.3.1.3 Khmer

Khmer (Cambodia; Austro-Asiatic) is a language classified as mixed with regard to CNNCs. Modern Khmer has two alternating patterns of CNNCs with neither primary, namely {N,NUM} and {N,NUM,CLF}, as shown in (8.13). In (8.13) the numeral classifier *néak* (lit. ‘person’) is optional (hence in parentheses). The pattern of {N,NUM} is generally used in spoken Khmer, whereas the pattern of {N,NUM,CLF} is normally used in written Khmer or in careful speech (Jacob 1965).

(8.13) *Khmer (Ancient and Contemporary) (Jacob 1965: 145)*

<i>ko:n</i>	<i>pi:</i>	<i>(néak)</i>
<i>child</i>	<i>two</i>	<i>(CLF)</i>
<i>‘two children’</i>		

According to the earliest Khmer inscriptions (7th-12th centuries), the two patterns of CNNCs were already recorded as shown in (8.14 a-b).

(8.14) *Khmer ((Pre-Angkor inscription, 7th-12th centuries) (Jacob 1965: 152)*

(a)

<i>kon</i>	<i>2²⁷</i>
<i>child</i>	<i>two</i>
<i>‘two children’</i>	

(b)

<i>ton</i>	<i>mvay</i>	<i>tem</i>
<i>coconut</i>	<i>one</i>	<i>CLF (lit. ‘tree’)</i>
<i>‘one coconut tree’</i>		

²⁷ In the pre-Angkor inscription, numerals are usually written in figures (Jacob 1965: 152).

Previous studies (Jones 1970; Huffman 1973; Adams 1991) suggest that the pattern of {N,NUM,CLF} is not native to Khmer and other Austro-Asiatic languages. The development of the numeral classifier systems in the family have been attributed to long contact with numeral classifier languages in the area, typically Tai-Kadai and Sino-Tibetan languages (where numeral classifiers are obligatory). This is supported by the fact that numeral classifier systems in many Austro-Asiatic languages including Khmer are optional; certain languages in the family do not even possess a numeral classifier system.

The inconsistency and absence of numeral classifier systems in Austro-Asiatic as mentioned above are confirmed by Gil's (2005) sampling. Among 14 Austro-Asiatic languages investigated, only 6 show obligatory systems of numeral classifiers, which in the remaining languages are either optional or absent, especially in the far-flung genera like Munda, mainly spoken in central and eastern India. Munda represents the primary split of Austro-Asiatic before the Austro-Asiatic speakers spread to Southeast-Asia (Anderson 2006: 598). As it is hypothesised that numeral classifiers in Austro-Asiatic are attributed to the close relationship with Southeast Asian languages, therefore special attention regarding numeral classifiers should be paid to the Munda languages.

According to the specimens taken from the same folk tale by Grierson (1904) in 15 Munda languages²⁸ where CNNCs are available, only Korwa shows the use of numeral classifiers, and their use is optional. Some of these other languages may possess numeral classifier systems but the systems were not recorded in the specimens (e.g. Santali and Kharia; Gil 2005). Also it is possible that numeral classifiers may not be employed with lower numerals which are mostly indigenous numerals; rather they may be used with the loan Indo-Aryan numerals which are not present in the specimens. Below are examples expressing 'one goat' from some of the specimens.

²⁸ They include Kherwari, Santali, Mundari, Bhumij, Birhar, Koda, Ho, Turi, Asuri, Korwa, Korku, Kharia, Juang, Savara and Gadaba.

<i>Birha:</i>	<i>mia</i>	<i>bhedi (one-goat) (p.105)</i>
<i>Korwa</i>	<i>mi(t')-got</i>	<i>pathru (one-piece goat) (p.152)</i>
<i>Korku</i>	<i>mia</i>	<i>shiri (one-goat) (p.175)</i>
<i>Kharia</i>	<i>moi</i>	<i>merom (one-goat) (p.202)</i>
<i>Juang</i>	<i>min</i>	<i>mera (one-goat) (p.214)</i>

It should be noted that the numeral classifier systems in Munda languages might have been influenced mainly by Indo-Aryan languages due to the fact that the numeral classifiers are used only with the Indo-Aryan numerals, especially the higher numerals (Grierson 1904; Emeneau 1956: 14). Therefore, with the comparative method for reconstruction (assuming that the inherited characteristics from the protolanguage are expected to be present in most descendant languages) and the distribution of numeral classifier languages in the family, it seems likely that {N,NUM,CLF} did not exist in Proto-Austro-Asiatic. The statement made in the prior research is tentative, however, so there is no concrete evidence yet suggesting the existence of classifier systems in Proto-Austro-Asiatic.

Although it is not stated explicitly in the prior research what the original state of CNNCs in Khmer (or in Austro-Asiatic languages) was like, due to the optionality of the presence of classifiers in CNNCs in most genera, the pattern of {N,NUM} is conjectured to be the native pattern of Austro-Asiatic including Khmer. This may be accounted for by the fact that in addition to {N,NUM,CLF}, the pattern of {N,NUM} is another pattern of CNNCs used in Khmer since the earliest inscriptions, for example, *kon 2* 'two children', *krapi 4* 'four buffaloes' (Jacob 1965: 152). It seems likely that the pattern of {N,NUM} is the pattern potentially present in most genera in the Austro-Asiatic family, at least as an alternating pattern if not the primary one. The pattern can be observed more frequently, particularly in Austro-Asiatic languages spoken in India, where the patterns of {N,NUM,CLF} and {N,NUM,NSG} are predominant.

However, it is not yet clear how and why Khmer developed the pattern of {N,NUM,CLF} if it already had the simple {N,NUM}. Considering the phenomenon that numeral classifiers spread from one language to another via the borrowing of numerals (cf. Japanese in §8.3.1.2), Khmer or the pre-Khmer language might be understood accordingly. It is noted in Haarmann (1990: 84) that the Khmer numerals 30-100 are borrowed from Thai. It is possible that this is the reason the numeral classifiers spread from Thai to Khmer along with the numerals.

8.3.2 $\{N,NUM,CLF\} < \{N,NUM,CLF,SG\}$

Kana (Niger-Congo; Nigeria), a Cross River language, has {N,NUM,CLF} as a dominant mode of CNNCs. For example,

(8.15) *Kana* (Ikoro 1994: 17)

<i>zii</i>	<i>kā</i>	<i>wā</i>
<i>one</i>	<i>CLF</i>	<i>wife</i>
<i>'one wife'</i>		

According to Ikoro (1994), Proto-Cross River presumably had noun classes. This is because most languages in the genus show traces of noun classes which are fused with number in one morpheme (albeit the noun class in many cases may serve no more function). Efik, another Cross-River language, still shows some of the few singular/plural alternations as shown in the examples below. Accordingly, pre-Kana must therefore have had {N,NUM,SG} and {N,NUM,NSG} for CNNC_{SG} and CNNC_{NSG} respectively.

<i>Singular</i>	<i>Plural</i>	
<i>o-fin</i>	<i>i-fin</i>	<i>'slave'</i>
<i>ɔ-bɔŋ</i>	<i>n-bɔŋ</i>	<i>'king'</i>
<i>e-dên</i>	<i>í-dên</i>	<i>'male'</i> (Cook 1969a in Ikoro 1994: 9)

Since it seems unlikely that {N,NUM,CLF} changed directly from {N,NUM,SG}, however, {N,NUM,CLF} may rather have developed from an intermediate construction, such as {N,NUM,CLF,SG}. This conjecture is supported by the current patterns in Ejagham (for Ejagham, see §8.4 below) and the Southern Bantoid

languages where {N,NUM,CLF,SG} and {N,NUM,SG} co-exist. The co-existence of the two patterns may suggest the initial stage of the development of a numeral classifier system in a language with noun classes. After the loss of the noun class system in the language, the numeral classifier systems may then be fully developed. Ikoro (1994: 24) conjectures that

The number of numeral classifiers in Ejagham is small when compared to Kana. One may assume that this is due to the fact that there is also a functioning noun class system.

In other words, the loss of the noun class system in Kana may account for the expansion of the numeral classifier system in Kana. This idea is not implausible when considering the complementary distribution of the systems of plural marking and numeral classifiers. Kana has lost noun classes and it has consequently lost the number distinction in nouns. The loss of number distinction makes the nouns in the language vague in number. It seems likely that non-number marking languages may adopt numeral classifiers more easily than number marking languages. This is because the nouns in non-number marking languages may be easily treated as collective-like nouns and then may require numeral classifiers. Kana may belong to such a case. On the contrary, in Ejagham the noun class system is still functioning, and so is the grammatical number. Therefore, most nouns are clearly singular and thus can be counted without unit counters such as a numeral classifier. Even though there is a subsystem of numeral classifiers in Ejagham, the number system hidden in the noun class system may block the growth of numeral classifiers.

In any case, the motivations for the rise of {N,NUM,CLF} in Kana (and its sister languages) remain obscure. Ikoro hypothesises that {N,NUM,CLF} might have been inherited from Proto-Kegboid, the immediate parent language of Kana. This is because some other Kegboid languages such as Baan and Gokana spoken in Nigeria also have numeral classifier systems (Ikoro 1994: 24)

Then the question becomes how the Proto-Kegboid developed the numeral classifier system. Ikoro (1994: 24-25) claims that the numeral classifier system in Proto-Kegboid might have been influenced by the numeral classifier system in Ejagham

because the Ejagham area might have been the homeland of the Proto-Kegboid language. However, Ikoró touches on this issue very briefly and there is no concrete evidence supporting his claim. If indeed Kana borrowed the system of numeral classifiers from Ejagham via language contact, the remaining question would be how the system spread. Would it spread via numerals? Since there is no evidence available, the development of the numeral classifier system in Kana through language contact will be left open at this stage.

Considering the development of the numeral classifier systems in Chinese and Ejagham (cf. §8.4), we may conjecture that the system starts with some discrete spatial nouns (which are culturally perceived as uncountable or can be counted in various ways), and later that it has spread to other nouns in the language. Kana might have taken the same route; however, there is no evidence for this argument.

Another possibility worth mentioning is that Kana might have developed the numeral classifiers from *noun classifiers* (cf. §5.2.3.2) or *noun class markers* (cf. §5.2.3.3). This idea comes from the observation given in Ikoró (1994) that numeral classifiers in Kana form a morphological unit with the nouns rather than with numerals. For example, the pro-clitic /i/ for the diminutive (i.e. a morpheme denoting *small size*) is normally attached to the noun, but when the diminutive is used in CNNCs, it is attached to the numeral classifier instead, as shown in the example (8.16) below.

(8.16) *Kana* (Ikoró 1994: 21)

zù	í	kà	núú
<i>one</i>	<i>DIM</i>	<i>CLF</i>	<i>rat</i>
<i>'one small rat'</i>			

This suggests that numeral classifier and the noun form a morphological unit, and hence that the other element is not allowed to be inserted between them. Another piece of evidence is that an adjective is allowed to be inserted between a numeral and a numeral classifier which appears next to the noun, as illustrated in the example (8.17) below.

(8.17) *Kana* (Ikoro 1994: 22)

zìì pià kà péé
one bad CLF goat
'one bad goat'

Therefore, considering the fact that noun classifiers or noun class markers are attached to nouns and that they are in fact subtypes of the nominal categorization systems (Grinevald 2002a), the idea that the origins of numeral classifier systems in Kana may be associated with noun classifiers or noun class markers is plausible.

8.4 {N,NUM,CLF,SG}

{N,NUM,CLF,SG} < {N,NUM,SG}

This developmental pattern indicates that {N,NUM,CLF,SG} may develop from {N,NUM,SG}. This is illustrated by Ejagham (Niger-Congo; Cameroon, Nigeria). It would be interesting to know how the numeral classifier system in Ejagham arose among the languages with well-preserved noun class. There are two current patterns employed in Ejagham at least, namely {N,NUM,SG} and {N,NUM,CLF,SG}.

(8.18) *Ejagham* (Watters 1981: 469)

a. {N,NUM,SG}

N-díg *má-d*
CL-rope *CL-one*
 'one rope' (CL *N-* = singular number)

b. {N,NUM,CLF,SG}

è-rám ' *í-čókùd já-d*
CL-CLF *GEN*²⁹ *CL-orange* *CL-one*
 'one orange' (CL *è-* = singular number)

There is no historical evidence showing a change from {N,NUM,SG} to {N,NUM,CLF,SG}. However, two observations about numeral classifiers in this language suggest that the numeral classifier constructions are not fully developed. Therefore, we may conclude that {N,NUM,CLF,SG} is relatively innovative, compared to {N,NUM,SG}. The first observation is that, generally, in a numeral classifier language, most nouns are assigned numeral classifiers. In Ejagham, however, the numeral classifiers have not yet spread to other nouns, rather being mostly limited to nouns denoting *plants* or *trees*, as noted in Watters (1981: 313).

²⁹ Aikhenvald (2000: 99) calls this kind of tone as a 'genitive linker'.

Most nouns do not use a classifier when being enumerated, but for the various types of nouns [...], e.g. seeds, grains, kernels, nuts, long fruits, and roots, round fruits and roots, plants, trees, and vegetables, a classifier is generally required.

These nouns can be counted in various ways. For example, the noun ‘orange’ can be counted in terms of *tree* or *fruit*. Therefore, numeral classifiers are required to help clarify what exactly is being counted.

The second observation is that numeral classifiers in a numeral classifier language have the status of a lexico-grammatical morpheme (Grinevald 2002a: 260). But in this language, its status is similar to that of lexical nouns rather than that of grammatical elements. This can be seen from the fact that the numeral classifiers can be marked for noun class and number and can even govern the other elements in the noun phrase construction. In other words, they can be treated as the head of the noun phrase construction. It can be seen from the example that the head precedes the modifier. This may trigger *the two-head construction*, i.e. a noun phrase construction with two head nouns (cf. Simpson 2005). In this case, one is the numeral classifier and the other is the noun. And between the two heads, the numeral classifier which is semantically more generic becomes the left most head as shown below.

(8.19) *Ejagham* (Watters 1981: 469)

<i>è-rám</i>		<i>í-čókùd</i>	<i>jǎ-d</i>
CL-CLF	GEN	CL-orange	CL-one
Head 1 st		Head 2 nd	Modifier
‘one orange’			

The numeral classifier *rám* is used to denote any fruit (which is round) (Watters 1981: 311). So the numeral classifier is more generic than a specific noun like *í-čókùd* ‘orange’ (cf. *orange tree* in English where the head is *tree* and the modifier is *orange*). Since the noun *orange* behaves like a modifier for the noun ‘tree’ and the word order in *Ejagham* places the head before the modifier, the numeral classifier *rám* ‘round fruits’ precedes the noun *í-čókùd* ‘orange’. Then both the numeral classifier and the noun in turn precede the numeral *jǎ-d* ‘one’.

In sum, the motivation for change is similar to Chinese. That is to say, some discrete spatial nouns in the language can be counted in various ways, and so the language requires lexical nouns for clarity as far as counting is concerned. The functions of these nouns are not different from numeral classifiers. The difference is that they remain more lexical.

9 Historical Origins of $CNNC_{NSG}$

This chapter explores the historical development of the various structural patterns of current $CNNC_{NSG}$ as illustrated in Chapter 6. The types which will be explored are {N,NUM} (§9.1), {N,NUM,NSG} (§9.2), {N,NUM,CLF} (§9.3), {N,NUM,CLF,NSG} (§9.4), {N,NUM,OBL,SG} (§9.5), and {N,NUM,OBL,NSG} (§9.6).

9.1 {N,NUM}

The {N,NUM} pattern is the simplest pattern of $CNNC_{NSG}$, ignoring the pattern of {N,NSG} (referring to {N,DU} or {N,TRI}) where the non-singular marker is not a cardinal numeral proper). There are a couple of sources for the pattern of {N,NUM}, namely {N+ ‘two-ness’} (§9.1.1), {N,NUM,NSG} (§9.1.2), and {N,NUM,CLF} (§9.1.3).

9.1.1 {N,NUM} < N + ‘two-ness’

This historical pattern suggests that the pattern of {N,NUM} (where Num refers to the numeral ‘two’ only) may develop from a syntactic construction in which the noun became modified by the words meaning ‘two-ness’, such as ‘company’ and the like, through the process of numeralization (cf. §4.4). This historical pattern is supported by diachronic and synchronic evidence. For example, the words meaning ‘wings’ or ‘eyes’ (or things that come in twos) may be used to denote the number ‘two’ (Menninger 1969: 119). For example, in Kwazá (Kwaza; Brazil), the word *aky* means ‘to be two’ and ‘company’ (Voort 2004: 214). According to the general tendency of numeralization, it seems likely that the word *aky* originally had the meaning ‘company’, and has later come to be also used as the numeral ‘two’, rather than the other way round. In Eskimo-Aleut, the numeral ‘two’ has the stem meaning ‘one that follows’ or ‘wave’ (Bonnerjea 1978: 53). In Wari’ (Chapacura-Wanhan; Brazil), native numerals are not available, but the verb *tucu caracan* ‘to face each other’ denoting the number ‘two’ can be used (although nowadays it is actually rarely used) (Everett and Kern 1997: 347-348). The

issue of words with a numerical interpretation has already been mentioned in § 4.4, so it will not be repeated here. The stage of $N+$ ‘two-ness’ can be regarded as an embryonic stage before the rise of $\{N,NUM\}$.

As in the case of $CNNC_{SG}$, this change is perhaps attributed to the user’s need to express the exact number (i.e. *two*). That is, the change involves the principle of expressivity and economy (cf. §8.1.1). However, the principle of economy also plays a role in this matter. Instead of creating a new linguistic symbol, humans may prefer to make use of internal sources to refer to the similar concept. Metaphorical extension such as this is a frequently observed phenomenon, giving rise to the polysemy of words. In sum, the overall change deals with semantico-syntactic change (i.e. a change involving both semantic and syntactic aspects), as the semantic change makes syntactic constructions like $N+$ ‘two-ness’ turn into $\{N,NUM\}$.

9.1.2 $\{N,NUM\} < \{N,NUM,NSG\}$

The historical pattern of $\{N,NUM\} < \{N,NUM,NSG\}$ suggests that $\{N,NUM\}$ may arise from $\{N,NUM,NSG\}$. This pattern is evidenced by Celtic languages (Irish and Welsh), Creoles, and Arabana-Wangkangurru (Pama-Nyungan; South Australia). These three examples illustrate different contributory change factors.

9.1.2.1 Celtic (Irish and Welsh)

In Old Irish (8th-9th centuries), nouns with numerals from *three* to *ten* stand in the plural (*two* required the dual form), as shown in (9.1) (McCone 2005: 62). However, in Modern Irish (13th century- present), the numerals from *three* to *ten* are followed by either a singular or a plural, as shown in (9.2) (Ò Dochartaigh 1992: 55).

(9.1) *Old Irish (William Gillies, personal communication.)*

<i>trí</i>	<i>caitt</i>
<i>three</i>	<i>cat.PL</i>
	<i>‘three cats’</i>

(9.2) *Modern Irish (William Gillies, personal communication.)*

a. *trí* *cait*
 three *cat.PL*
 ‘*three cats*’

b. *trí* *chat*
 three *cat*
 ‘*three cats*’

The structural change in this case is caused by *a change of the phonological system* of the language, namely, the phonetic erosion of final unaccented syllables. This leads to the loss of some case and number marking, leaving the plural forms formally indistinguishable from their singulars. To clarify nominal number, the language developed new number markers. However, in the Irish dialects, the number markers were not introduced in CNNC_{NSG} with numerals greater than 2 where the number of entities is made clear by the numeral, and so the qualified nouns can stand in the singular (Acquaviva 2006: 1866).

The situation in Welsh is quite similar to Irish. In Welsh, the loss of number in quantificational expressions is perhaps owing to drastic sound changes such as happened in the Irish language. It is generally thought that heavy stress in Late British (middle of the fifth century until the earlier half of the sixth century) resulted in the loss of final syllables, where grammatical number and case markers are supposed to have been present. This change also gave rise to Neo-British, namely Welsh, Cornish, and Breton. This process was completed around the middle of the sixth century (Jackson 1953: 5, 618; Watkins 1993: 289).

9.1.2.2 Creoles and Pidgins

Generally speaking, pidgins are simplified languages which are developed as a medium of communication between groups of people who need a common language. Such a group of languages evolved through trade or other contacts between the native inhabitants (typically the inhabitants of countries in the Pacific and the Atlantic) and

Europeans during the European colonization between the 17th and the 19th centuries. Pidgins are a combination of the vocabulary of the *superstrates* or *base languages* (i.e. the languages of the new settlers who have more power, typically Europeans such as the English, French or Portuguese) and the grammatical structures of the substrates (i.e. the languages of the native inhabitants). The striking feature of pidgins is their simplified grammatical structures. For example, the morphological inflections typical of European languages (such as number and gender markers) are removed. In other words, most of the pidgins are simplified versions of one of the European languages. A pidgin is like a language created ad hoc and so it does not have native speakers and the range of its use is limited. Later when a pidgin is increasingly learned and spoken *natively* by generations in a whole community and its grammatical structure and vocabulary are elaborated, the pidgin may become a *creole*. So a creole is, in other words, a nativized pidgin of the speech community. It is used for all purposes of the speech community. Angolar, spoken in São Tomé Island, western Africa is an example. It is a Portuguese-based creole with a substrate of Angolan Bantu languages, spoken by native African labourers and slaves from Angola (Matthews 1997: 81-82; Holm 2000: 5-6; Romaine 2006: 600-601; Anderson et al. 2006: 751-752).

This section suggests an evolutionary path for CNNC_{NSG} in Creoles and Pidgins (CP) based on a sample of 10 languages representing 5 different macro areas. The evolutionary path in CP is different from those of other languages or language groups. That is to say, in CP, the evolutionary path starts from the current CNNC_{NSG} of the superstrates, i.e. European and Arabic languages, which use the pattern of {N,NUM,NSG} for their CNNC_{NSG}. Therefore, the evolutionary paths of CNNC_{NSG} in CP refer to the structural change from the superstrates to the CP; not within the creole or pidgin language itself.

There are 2 structural types found in CP, namely {N,NUM,NSG} and {N,NUM}. Almost all the CP in my sample show a historical development from {N,NUM,NSG} to {N,NUM}. This is because of the tendency of CP to simplify the superstrate (which may be rich in grammatical categories especially in the case of

European languages) by avoiding redundant complications such as inflections including grammatical number markers.

The prominent characteristic of the historical development of $CNNC_{NSG}$ in CP is that almost all CP languages take the same route, even though they are not genetically or geographically related. The immediate question then arises: why is there such a similarity among CP languages?

According to Holm (2000: 28), there are currently two schools of thought on the origin of creoles. One is the *universalist* and the other is the *substratist*. In the universalist theory, led by Adolpho Coelho, the nineteenth century Portuguese philologist (Holm 2000: 27), the form of creoles (e.g. simplification) should be attributed to “certain universal tendencies in second language learning by adults rather than to the influence of substrate languages” (Holm 2000: 27-28).

Regarding the substratist suggestion, proposed by Lucien Adam, the French philologist (1883 in Holm 2000: 28), creoles are influenced by the substrate, as illustrated by an Atlantic Creole and various African languages, where the formation of the plural with the third person plural pronoun and some phonological features can be attributed to the substrate.

It may be the case that both theories are right to some extent, and it may also depend on which aspects of grammar are under discussion. Regarding $CNNC_{NSG}$, if the extra elements are number markers, they seem to be always dropped despite the fact that the $CNNC_{NSG}$ in the substrates is also {N,NUM,NSG}. So this phenomenon is very consistent with a universalist view. However, there is evidence that if the extra elements are numeral classifiers, they might be used in CNNCs. This is illustrated by Chinese Pidgin English below. Thus the substratists are probably right for this case.

(9.3) *Chinese Pidgin English (Holm 1988: 516)*

trí pisi tébəl
three CLF 'lit. piece' table
'three tables'

So far when compared with other groups of languages, it can be noticed that the CP group fits well with the general principles of language evolution, namely the principles of economy and distinctness. Therefore, the pattern of {N,NUM} is suitable for serving a communicative function. Also, CP reflect the unique development of CNNC_{NSG}. That is, regarding CNNC_{NSG}, many languages are likely to develop towards complexity rather than simplicity. On the contrary, for creoles, the CNNC_{NSG} develops towards simplicity.

9.1.3 {N,NUM} < {N,NUM,CLF}

This pattern indicates that {N,NUM.CLF} or {N,NUM}, as illustrated in (9.4), is derived from {N,NUM,CLF}.

- (9.4) *Nivkh (Gruzdeva 1998: 62)*
- | | |
|------------------|----------------|
| <i>n'ivŋ</i> | <i>menŋ</i> |
| <i>man</i> | <i>two.CLF</i> |
| <i>'two men'</i> | |

As already mentioned (§8.1.2), {N,NUM.CLF} is a simplex construction where the concepts and forms of a numeral and a numeral classifier are fused into one morpheme. {N,NUM}, which arises from {N,NUM,CLF}, has already been described in §8.1.2, but in that section the focus was on CNNC_{SG} and attention was concentrated on only one language, namely Beijing Mandarin. In this section, we will look at CNNC_{NSG}. Once again, Beijing Mandarin provides evidence of this historical pattern. According to Chirkova (2004: 1), the fusing of the numeral *one* with the numeral classifier *ge51* has also spread to the numerals *two* to *ten* in spoken Beijing Mandarin, particularly in the case of the numerals *liǎng* 'two' and *sān* 'three', which become *liǎ* and *sā* respectively. Both of these forms can be used attributively and nominally. They are regarded as fully lexicalized; that is, the two new forms *liǎ* and *sā* express two new meanings, namely 'two items of' and 'three items of' respectively. Both *liǎ* and *sā* have their own character transcriptions and have also been included in most standard dictionaries. Although the numerals four to ten have also been fused with *ge51*, they are rarely used and are not yet

lexicalized, so they do not have character transcriptions and are accordingly not included in dictionaries.

Unfortunately, the process of change is not mentioned at all in Chirkova (2004). Since there is no available historical evidence which shows the process of change from {N,NUM,CLF} to {N,NUM} *within* a language, we may need to view this developmental pattern *across* languages, based on the assumption that the diversity of structural patterns mirrors the (hypothetical) evolutionary ladder. Four languages are selected to represent the process of change, namely Semelai (Austro-Asiatic; Malaysia), Taba (Austronesian; Indonesia), Warekena (Arawakan; Brazil and neighbours), and Nivkh (isolate; Siberia, Russia). These languages display a typological continuum and a diachronic pattern of language change from {N,NUM,CLF} to {N,NUM} (where the numeral classifier is fused with the numeral), as illustrated in Figure 9.1 below.

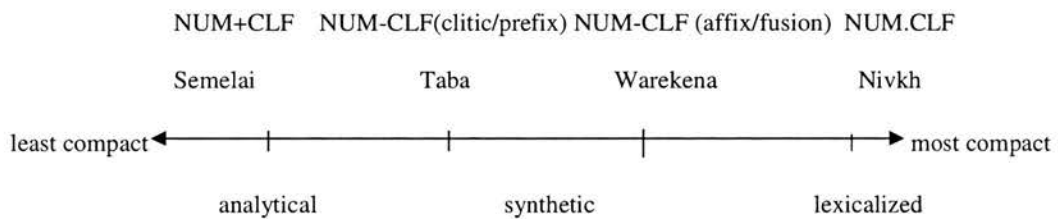


Fig. 9.1 Continuum of compactness of numeral plus numeral classifier

Figure 9.1 illustrates various degrees of compactness in the combination of numerals and numeral classifiers from language to language. The least compact language in terms of numeral and numeral classifier is found on the analytical end (Semelai, exemplified in (9.5) below), where both numerals and numeral classifiers are free forms.

- (9.5) *Semelai* (Kruspe 2004: 204)
- | | | |
|----------------------|-------------|-------------|
| <i>tmpɔh</i> | <i>bjɛʔ</i> | <i>cʰɔŋ</i> |
| <i>seven</i> | <i>CLF</i> | <i>hill</i> |
| <i>'seven hills'</i> | | |

9.2 {N,NUM,NSG}

{N,NUM,NSG} is one of the most interesting cases. Here the non-singular marker is seemingly redundant when used in CNNCs. The immediate question thus arises: why is the non-singular marker required in the construction? There are a few historical routes which may have led to the current pattern of {N,NUM,NSG}.

9.2.1 {N,NUM,NSG} < {N,NSG} where NSG is plural

One possible route is for {N,NUM,NSG} to have developed from {N,NSG}. The non-singular markers here refer to plural markers only (for dual marker, see §9.2.2 below). As there is no direct evidence for this developmental pattern, the argument is based mainly on grammaticalization theory. The pattern of {N,NUM,NSG} consists of three constituents, namely N, Num, and NSG. It is reasonable to assume that the three constituents were not originally combined together simultaneously. So, there are 3 logically possible combinations regarding the emergence of {N,NUM,NSG}.

- 1) The construction of {N,NUM} emerged first and then the non-singular was combined with it later. This possibility is represented by the notation: [[N,NUM]+NSG]].
- 2) The construction of {NSG,NUM} emerged first and then the noun was combined later, represented by the notation: [[NSG,NUM]+N]].
- 3) The construction of {N,NSG} emerged first and then the numeral was combined later, represented by the notation: [[N,NSG]+NUM]].

It is proposed that the third possibility is the most plausible, whereas the other two seem less likely. For the first possibility, it would appear to be quite difficult to find the speaker's motivation for adding the non-singular marker to {N,NUM} if the pattern of {N,NUM} being used is already sufficient in terms of practicalities. No languages in the current sample show this route (except the creoles, see §9.2.3). Regarding the second logical possibility, the combination of numeral and number does not make sense at all. It

may also be noticed that languages in which a plural marker is attached to a numeral are extremely rare. Wolof (Niger-Congo; Gambia and Senegal) is the only language in the current sample illustrating this, as shown below:

- (9.8) *Wolof (Ngom 2003: 48)*
ñaar-i *xarit*
two-PL *friend*
'two friends'

One possible explanation is that Wolof is a language where number distinction is expressed through postnominal determiners (cf. Ngom 2003: 19-33) as shown below:

- (9.9) *Wolof (Ngom 2003: 19)*
xale *bi/yi*
child *DEF.SG/DEF.PL*
'the child/children'

It is possible that numerals may be treated in the same way as determiners. It is not surprising that number distinctions can be made on some other elements such as verbs (i.e. verbal number) or determiners. However, although this is sensible for all determiners, it is not for the case of numerals, because the number of the referent is clear from the numerals. So it may be the case that Wolof, a language with noun classes, may lack or might have lost class-number agreement attached to nouns but the class-number system remains and appears on the modifiers. Then the class-number agreement becomes attached to modifiers (including numerals) (cf. Ngom 2003: 19).

The third possibility is, however, the most likely because it can be explained in terms of grammaticalization. Since a discussion of the third possibility overlaps with the typological continuum of obligatoriness of plural marking which was dealt with in §5.1.3 and also reflects the diachronic pathway of the pattern of {N,NUM,NSG}, this topic will not be repeated here.

9.2.2 {N,NUM,NSG} < {N,NSG} where NSG is dual

This pattern indicates that {N,NSG} in a language can develop into {N,NUM,NSG}. Due to the fact that the NSG in {N,NSG} to be discussed here is the dual marker only, the notation will be changed to {N,DU}. Also the NSG in {N,NUM,NSG} deals with the plural only, so the notation will be changed to {N,NUM,PL} when mentioned. Arabic (Afro-Asiatic) and Futuna-Aniwa (Austronesian; Vanuatu) provide helpful evidence for this developmental pattern.

9.2.2.1 Arabic

Based on the diachronic discussion of the split of the old dual morpheme from Classical Arabic to the Arabic dialects provided by Blanc (1970), we may summarize the change in patterns of CNNCs in Arabic as follows.

In Classical Arabic (8th-10th centuries), the pattern of {N,DU} was used to convey the two-ness of the referents. At that time, the dual marker functioned as a grammatical concord category. For example, the sentence *'inna hāḍayni lwaladayni kânā ṣadīqayni hamīmayn* 'these two boys were close friends' "or rendering dual forms by a subscript 2, 'these₂ boy₂ were₂ close₂ friend₂'" (Blanc 1970: 43). It can be seen that the dual is not suffixed only to the noun, but also to the other elements in the sentence. So the pattern with two referents can be represented as {N,DU} as in (9.10) below.

(9.10) *Classical Arabic (Blanc 1970: 43)*
lwaladayni
boy.DU
'two boys'

Later, in the Arabic dialects (which have developed from Classical Arabic), the dual is no longer a grammatical concord category. Instead, basically only nouns are marked for the dual, while other elements are not. Since the dual is used only with nouns, its function is not much different to that of the numeral *two*. That is, the dual only indicates the number of the referents; it is not an agreement marker. Accordingly, when

it is clear from the context that the referents are two, then the dual is not necessary. In the Arabic dialects some nouns do not have the dual form, but use the numeral *two* instead, and the plural form of the noun can be used to indicate the number 'more than one' including 'two'. So at this stage, the pattern of {N,NUM,PL} can be used instead of {N,DU} as in (9.11).

(9.11) *An Arabic dialect (Blanc 1970: 44)*

itnēn 'asakir

two policeman.PL

'two policemen'

According to Blanc, in many Arabic dialects, the use of the dual is restricted to only a small set of nouns, or it is used with special meanings. For example, instead of indicating the exact number 'two', it can refer to a small quantity, as in '*iršēn* 'two or a couple of piastres'. However, in some dialects the dual remains productive, such as in those dialects of the Syro-Mesopotamian area. Generally speaking, it is possible to see the use of {N,NUM,PL} and {N,DU} as alternating constructions. However, it is worth mentioning that there is also the pattern of {N,NUM,DU}, i.e. where the dual suffix is added even if the numeral 'two' is present. This pattern indicates the special sense of 'just two', for example, *fīh ihlimalēn itnēn* 'there are (just) two possibilities.' (Blanc 1970: 44).

In sum, the historical change in the use of dual constructions suggests that the {N,DU} pattern may have split into three current patterns, namely {N,DU}, {N,NUM,DU} and {N,NUM,PL}.

9.2.2.2 Futuna-Aniwa

In Futuna-Aniwa (Austronesian; Vanuatu), there are two patterns for expressing the two-ness of things—namely {N,DU} and {N,NUM,DU}, as exemplified in (9.12a-b).

(9.12) *Futuna-Aniwa* (Dougherty 1983 in Dryer 1989b: 869)

a. {N,DU}

ru *fare*

DU *house*

'two houses'

b. {N,NUM,DU}

ru *tagata* *e* *rua*

DU *man* *two*

'two men' (NB: *e* = not glossed, but probably NUMPCL)

The examples in (9.12a-b) illustrate two patterns of CNNC_{NSG} (Dual only). One is the use of the dual marker alone, which Dryer (1989b: 869) refers to as the *dual word*, and the other is the use of the dual marker along with the numeral 'two'. According to Dryer (1989b: 869), this dual marker is cognate with the numeral 'two'. However, the dual marker 'ru' also has the grammatical function of an article because it is present in the article paradigm. Following a general tendency in many languages, it is possible that the dual marker has been grammaticalized from the numeral 'two'. This is because it shows grammatical meaning (the fact that it is used as an article) and also because it shows phonetic erosion (the loss of the final vowel). Once it assumed the function of an article, the meaning of the numeral 'two' was bleached and almost lost. That is the reason why the numeral 'two' was required again when the speakers wanted to express the two-ness of referents. So the development can be assumed to have started from {N,NUM} and then changed to {N,DU} and then {N,NUM,DU} via the grammaticalization process.

9.2.3 {N,NUM,NSG} < {N,NUM}

This developmental pattern indicates that {N,NUM} develops from {N,NUM,NSG}, as illustrated by Creoles and Pidgins (CP). Looking at data from Pidgins and Creoles, there are 4 out of 12 languages in the sample which adopt the {N,NUM,NSG} pattern (although somewhat incompletely) from the superstrate languages, in which the $CNNC_{NSG}$ requires a non-singular marker. This situation may be the consequence of *decreolization*, the historical process by which a creole develops towards the standard or prestige languages from which it derived. Decreolization is motivated by social reasons; that is, the speakers of the creole wish to standardize their language by adopting grammatical features of the superstrate language (which is regarded as more prestigious), and by dropping non-standard grammatical features (Holm 2000: 50).

As shown in the types and distribution of $CNNC_{NSG}$ in CP, there is some degree of variation in $CNNC_{NSG}$. In some languages, {N,NUM,NSG} used along side {N,NUM} as an exceptional case (e.g. Berbice Dutch Creole), an alternative mode (e.g. Belizean, Nigerian Pidgin, Nubi) and a primary mode (Hawaiian Creole). This suggests varying degrees of decreolization.

The change from {N,NUM} to {N,NUM,NSG} is rare, and evidenced only by creoles and pidgins. This is interesting because it suggests that in languages with {N,NUM,NSG}, the starting point is not likely to be {N,NUM}, but rather {N,NSG}. Cases where {N,NUM} have turned into {N,NUM,NSG} can be attributed to decreolization only. Decreolization also implies that conscious choice of the speakers is involved in the change. That is, when the speakers realize that the use of non-singular marker in $CNNC_{NSG}$ sounds standard, then the speakers try to change it. This suggests a unique kind of change which contradicts the principle of economy.

9.2.4 {N,NUM,NSG} < {N,NUM,OBL,NSG}

This pattern indicates that {N,NUM,NSG} may develop from {N,NUM,OBL,NSG}. It is illustrated by the history of English and Arabic. Although both represent the same developmental pathway, the factors of the change are different.

9.2.4.1 English

In Old English (mid-5th – 11th centuries), numerals can be used both adjectivally (alias *attributively*) and nominally. The numerals which can be used adjectivally are the numerals up to 19, and these behave just like most present-day English numerals; that is, they modify nouns and are placed directly before such nouns. The numerals which can be used nominally are high round numerals, typically multiples of ten such as 20 and 30. These behave like nouns, requiring the quantified nouns to be in the genitive plural (cf. §5.3.4). The numerals 100 and 1000 are both regularly followed by a noun in the genitive plural (Mustanoja 1960: 291; Marsden 2004: 380).

Based on several Old English grammars, two structural types of CNNC_{NSG} are observed, paralleled with the syntactic behaviours of numerals. One is {N,NUM,NSG}, where the numerals are used attributively and where nouns are basically in the nominative plural (when the noun phrase is in a nominative case position, i.e. normally unmarked) as in (9.13a-b). The other is {N,NUM,OBL,NSG} where the nouns are in the genitive plural, as in (9.13c). The latter type is probably more common, especially with high round numerals. As the inflectional endings began to disappear, the adjectival use of numerals (i.e. {N,NUM,NSG}) became more widely used, as seen in Middle English (1066 AD-1500 AD), shown in (9.14a), although some instances of {N,NUM,OBL,NSG} survived, as shown in (9.14b-c) (notice the use of *of*-periphrasis instead of genitive case in (9.14c)) (Mustanoja 1960: 291; Marsden 2004: 380; Mossé 1979: 52).

(9.13) *Old English*

- a. *twēgen* *scea þan*
 two *criminal.PL*
 ‘two criminals’ (Marsden 2004: 380)

- b. *ff* *menn*
 five *man.PL*
 ‘five men’ (Quirk 1957: 37)
- c. *feower* *hund* *wintra*
 four *hundred* *winter.GEN.PL* (lit. ‘of winters’)
 ‘four hundred winters’ (Marsden 2004: 380)
- (9.14) *Middle English* (Mustanoja 1960: 291)
- a. *an* *hundred knyghtes*
 one *hundred knight.PL*
 ‘one hundred knights’
- b. *fele* *hundred wintre*
 many *hundred winter.GEN.PL*
 ‘many hundred winters’
- c. *of* *ladies* *foure and twenty*
 of *ladies* *four and twenty*
 ‘twenty four ladies’

The change in the grammatical behaviour of the high round numerals affects the CNNCs. In Old English they are used in {N,NUM,OBL,NSG} (i.e. genitive plural and later with *of*-periphrasis), and then in Modern English they are normally used in {N,NUM,NSG}. Hence in Modern English the expression **thirty of years* is ungrammatical. It may be concluded that the CNNC_{NSG} in English has become simpler because of the change in grammatical behaviour of numerals (i.e. from noun-like to adjective-like). The numerals tend to behave in the same way, namely they are used adjectivally. This kind of change in the grammatical behaviour of high-valued numerals such as took place in the history of English may be a tendency found in European languages in general. In Old Irish, high-valued numerals, such as *cét* ‘hundred’ and *míle* ‘thousand’, required the nouns to be in the genitive plural (Thurneysen 1975: 244), but in Scottish Gaelic and Modern Irish (the descendants of Old Irish), these numerals are

followed by the singular instead (Gillies 1993: 181). Another example is Norwegian (Indo-European, Germanic), a language which developed from Old Norse. In Old Norse, the high-valued numerals are nouns and they are followed by nouns in genitive plural, for example, *fimm hundruð manna* [five hundreds of men] ‘five hundred men’. But in the Norwegian spoken today, the genitive case is dropped, giving rise to *fem hundre menn* [five hundreds men] (NB: hundred in plural) ‘five hundred men’ (Kinn 2004: 1).

In summary, this phenomenon is compatible with the universal proposed in Corbett (1978: 61), that “if numerals vary in behaviour then the higher will be nounier,” and so constructions which consist of high numerals require genitive insertion only to avoid double nominatives (cf. §5.3.4). The fact that {N,NUM,OBL,NSG} occurs for higher numbers in many other Indo-European languages, both contemporary and even older languages such as Latin (Clackson 2004: 805), suggests that the value of cardinal numerals in Indo-European varies in syntactic behaviour. The situation as such might have existed since Proto-Indo-European. The difference in grammatical behaviour between the high and low numerals suggests that the two numerals came into being at different times (see Hurford (1987 and 2001) for the grammatical idiosyncrasies of the low numerals). Over time, the high valued numbers have a general tendency to become more adjective-like, giving rise to the replacement of {N,NUM,OBL,NSG} with {N,NUM,NSG} as in English and Norwegian.

9.2.4.2 Arabic

In Old Arabic (i.e. Classical Arabic and Modern Standard Arabic or MSA³⁰) {N,NUM,OBL,NSG} for high numerals is employed as a means of forming CNNC_{NSG}, as shown in (9.15a-b). On the other hand, in New Arabic (New Arabic dialects), which has developed from Old Arabic, {N,NUM,NSG} is used, as shown in (9.16).

³⁰ It is noted in Procházka, (2006: 424) that MSA is practically identical in phonology, morphology, and syntax to CA, but it exhibits major differences from it in lexicon, phraseology, and style [...] The various dialects belong to a language type called ‘New Arabic,’ whereas both CA and (in spite of its label ‘modern’) present-day MSA are ‘Old Arabic.’

(9.15) *Old Arabic*

a. *MSA (Holes 1995: 173)*

ʔarbaát-u ʔawla:d-in
four.F-NOM boy.PL-GEN
'four boys'

b. *Classical Arabic (Benmamoun, 2003: 761)*

xamsat-u kutub-in
five-NOM book.PL-GEN
'five books'

(9.16) *New Arabic*

Maltese (Borg 1997: 268)

tlit bajdiet
three egg.PL
'three eggs'

Therefore we can see that {N,NUM,OBL,NSG} has changed into {N,NUM,NSG}. The change obviously involves the loss of oblique case (here genitive case). It is generally recognised that the loss of cases and mood endings in New Arabic is a major characteristic which differentiates it from Old Arabic. According to the evidence of old inscriptions, the loss of case endings might have taken place as early as the 1st century C.E. (Procházka 2006: 424). So Arabic is another case showing that the change in CNNCs is affected by another area of grammar.

9.3 {N,NUM,CLF}

The pattern of {N,NUM,CLF} can be derived from {N,NUM} and {N,NUM,NSG}. As the developmental pattern {N,NUM,CLF} < {N,NUM} has already been described in CNNC_{SG} in §8.3.1, it will not be repeated here.

{N,NUM,CLF} < {N,NUM,NSG}

This developmental pattern indicates that {N,NUM,CLF} can arise from {N,NUM,NSG}. This can be illustrated by Bengali (alias Bangla), an Indo-Iranian language spoken in eastern India. According to Dasgupta (2003), contemporary Bengali possesses {N,NUM,CLF}, as shown in (9.17) below.

(9.17) *Bengali (DaSGupta 2003: 380)*

<i>duṭo</i>	<i>boi</i>
<i>two.CLF</i>	<i>book</i>
<i>'two books'</i>	

This language developed from the Sanskrit language, an ancient Indian language. Following the traditional view for Indo-European, Sanskrit does not possess numeral classifiers, but rather uses the pattern of {N,NUM,NSG} instead. So it is reasonable to conclude that {N,NUM,CLF} is a new pattern in Bengali, and this is true at least when considering the line from Sanskrit or pre-Bengali to modern Bengali.

Chatterji (1926: 777-781) surveys some numeral classifiers in Bengali by comparing the forms with its sister languages and comes to the conclusion that the numeral classifier system originated in pre-Bengali and other sister languages such as Oriya, Assamese, and Maithili, no later than the New Indo-Aryan period or at least before 1000 AD.

Regarding the motivation for the emergence of the numeral classifier system in Bengali or other Indian languages, most scholars, among them, Emeneau (1956: 16), speculates that the numeral classifier systems spread from Southeast Asia to India. The evidence, however, remains inconclusive.

9.4 {N,NUM,CLF,NSG}

The pattern of {N,NUM,CLF,NSG} (which is a complex construction) may have developed from the patterns of {N,NUM,NSG} and {N,NUM}. The former is illustrated by Malto (Dravidian; India) (§9.4.1), and the latter by Chantyal (Sino-Tibetan; Nepal) (§9.4.2).

9.4.1 {N,NUM,CLF,NSG} < {N,NUM,NSG}

In Malto, there are two patterns of CNNC_{NSG}, namely {N,NUM,CLF} for non-human nouns³¹ and {N,NUM,CLF,NSG} for human nouns. These two patterns are shown in (9.18a) and (9.18b) respectively, but only the latter concerns us in this section.

(9.18) *Malto (Steever 1998: 372)*

- a. *tīni maq ōydu*
 three CLF cow
 ‘three cows’
- b. *tīni jen male-r*
 three CLF man-PL
 ‘three men’

It is conjectured that the numeral classifier systems are presumably not indigenous in the Dravidian languages, which include Malto. In the oldest Dravidian texts, which are written in Classical Tamil (150 BC- pre-fifth/sixth centuries AD) (Ronald Asher, personal communication; Rajam 1992), no numeral classifiers are observed. Rather, the patterns of {N,NUM} and {N,NUM,NSG} are already employed (Thomas Lehmann, personal communication). This is shown below.

³¹ Following the general tendency of the animacy hierarchy (cf. Haspelmath 2005: 142), in Malto, non-human nouns are not marked for number, whereas the human nouns are (Steever 1998: 362).

(9.19) *Classical Tamil*

a. *munuru* *ur*
three.hundred *town*
'three hundred towns' (Rajam 1992: 277)

b. *antaNar* *iruvar*
brahmin.PL *two.3PL*³²
'two brahmins' (Thomas Lehmann, personal communication)

In the Classical Tamil grammar written by Rajam (1992) (which is perhaps the most comprehensive among the Classical Tamil grammars (Ronald Asher, personal communication)), numeral classifiers are not mentioned at all. So the numeral classifiers might have been developed individually later. In addition, according to the distribution of numeral classifier systems in Dravidian languages, it is noticeable that numeral classifiers are not found in Southern Dravidian, for example Modern Tamil (Southern Dravidian, India, Sri Lanka) (see (9.20)). Rather they are observed in some languages of Central Dravidian, for example Kolami (Central Dravidian; Andhra Pradesh and adjacent areas) as in (9.21), and Northern Dravidian, for example Malto, as in (9.18) above—that is, the areas where the languages are in contact with Indo-Aryan languages.

(9.20) *Tamil (Lehmann 1989: 111)*

aintu nalla cattail-kal
five nice shirt-PL
'five nice shirts'

(9.21) *Kolami (Subrahmanyam 1998: 306).*

a. *{N,NUM,CLF,NSG} for Indo-Aryan numerals*
pāj jen mās-ur
five CLF man-PL
'five men'

³² Iruvar = a pronominal numeralized (cf. §8.1.2)

- b. {N,NUM,NSG} for native numerals
 ayd mās-ur
 five man-PL
 'five men'

So, viewing the change in CNNCs at the level of whole language family (i.e. from Proto-Dravidian or pre-Malto to modern Malto), we can see that {N,NUM,CLF,NSG} is relatively innovative, compared to {N,NUM} or {N,NUM,NSG}. Since there is evidence that the numeral classifier is used with Indo-Aryan numerals, but not native numerals, we may conclude that the emergence of {N,NUM,CLF} and {N,NUM,CLF,NSG} is attributable to language contact with Indo-Aryan languages. This observation is also made by Emeneau (1956: 14), as follows:

[...] Indo-Aryan classifier morphemes are used only with Indo-Aryan numerals in some of the non-Indo-Aryan languages [...] It spread thence to the other languages as a total construction consisting of numeral+classifier, and then was elaborated in some of the languages with native material, the native numerals, native morphemes as additional classifiers, etc.

9.4.2 {N,NUM,CLF,NSG} < {N,NUM}

Chantyal (Sino-Tibetan; Nepal) is categorized as mixed with respect to CNNC_{NSG}. There are four structural patterns used, namely {N,NUM}, {N,NUM,NSG}, {N,NUM,CLF}, and {N,NUM,NSG,CLF}. Only the last structural pattern concerns us in this section. This pattern, {N,NUM,CLF,NSG}, is only restrictively used: specifically, the numeral classifier is used with the numerals 1-3, and the non-singular marker is not obligatory. For example,

- (9.22) Chantyal (Noonan 2003: 318)
- | | | |
|-------------------------|---------------|----------------|
| <i>tin-ta</i> | <i>jəmməy</i> | <i>naku-ma</i> |
| <i>three-CLF</i> | <i>all</i> | <i>dog-PL</i> |
| <i>'all three dogs'</i> | | |

However, the question is what the original pattern in Chantyal or pre-Chantyal would have been, before the emergence of {N,NUM,CLF,NSG}. Old written records of Chantyal which might have shown the original CNNC_{NSG} pattern of the language are not available, but it is possible to make a deduction from (1) the patterns of CNNC_{NSG} in the Pre-Classical or Classical Tibetan texts (which are presumed to be the oldest texts in the Bodic languages, the genus to which Chantyal belongs) and (2) the patterns of CNNC_{NSG} in the Bodic languages themselves.

According to the two kinds of evidence, it is highly possible that {N,NUM} should be the pattern of proto-Bodic languages, and if this is proved true, it would suggest that other patterns are innovative. In Pre-Classical Tibetan and in Classical Tibetan, only one instance of {N,NUM} is found, as shown below, and the other patterns are not recorded at all.

(9.23) *Pre-Classical Tibetan (late 8th Century)*
deltar mchodrten bzhiyang brtsigslagsna
thus stupa four.TOP built.AUX.SUB
'Thus when the four stupas had been built...' (Denwood 1999: 269)

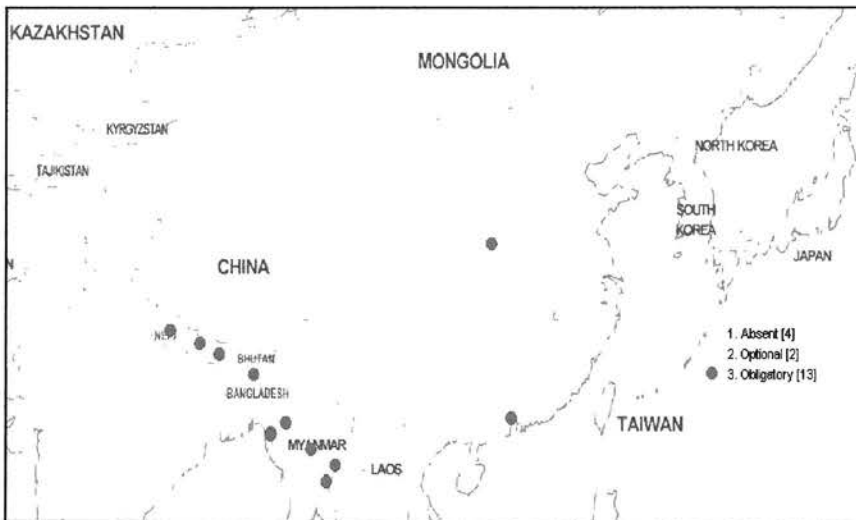
The five languages of the Bodic genus are listed below:

<i>Camling</i>	<i>{N,NUM,CLF}, {N,NUM,CLF,NSG}</i>
<i>Chantyal</i>	<i>{N,NUM}, {N,NUM,NSG}, {N,NUM,CLF,NSG}</i>
<i>Tamang</i>	<i>{N,NUM}</i>
<i>Tibetan</i>	<i>{N,NUM}</i>
<i>Nar-Phu</i>	<i>{N,NUM}</i>

It is noticeable that among the 5 languages, {N,NUM} is the most common. Although it would be quite risky to draw from this small sample the firm conclusion that {N,NUM} is the basic type of the proto-Bodic languages, it is nevertheless striking at least that {N,NUM} is found in 4 out of the 5 sampled languages. The most that can be said at this stage is that {N,NUM} seems to be the most likely candidate for the potential

basic type of Proto-Bodic, although further research would be required to confirm this hypothesis.

As generally recognized, {N,NUM,CLF} is commonly found in Sino-Tibetan languages. However, some scholars (see LaPollar 2003: 7, and references therein) argue that numeral classifiers do not exist in Proto-Sino-Tibetan (PST). There are two considerations which may make this view plausible. Firstly, geographically, the frequency of numeral classifiers is strikingly lower in the western part of this area, whereas in the eastern part almost all the languages have an obligatory classifier system. This is shown below in the map generated with WALS (Gil 2005).



Map 9.1 Geographical Distribution of numeral classifiers in Sino-Tibetan (Gil 2005)

According to the survey of numeral classifiers conducted by Gil (2005), among the 19 Sino-Tibetan languages, numeral classifiers in 13 languages are obligatory and in 4 languages are absent (all 4 are affiliated to the Bodic genus) and in 2 are optional. This suggests two possibilities. One is that the classifiers inherited from Proto-Sino-Tibetan are being lost in the western part of the family. Another is that numeral classifiers did not exist in Proto-Sino-Tibetan, but developed later in some individual languages (LaPollar 2003b: 27), especially the Sinitic languages. It is this second possibility which seems more plausible. This is because there is no remnant or trace of numeral classifiers

to be found in the non-numeral classifier languages in this family (cf. the remnant noun classes in Niger-Congo languages), so the suggestion that this system used to exist in the Proto-Bodic is not compelling. Also, the use of numeral classifiers in some Tibetan languages is not productive compared to the Sinitic languages. An example is Chantyal where the numeral classifiers are used with Nepali numerals. As pointed out in Noonan (2003), it seems evident that the numeral classifier system is not native to Chantyal, but it is borrowed imperfectly from its influential neighbouring language, Nepali (an Indo-Aryan language, spoken in Nepal), via the borrowing of the Nepali numerals. This claim is evidenced by the examples below. There are only two classifiers suffixed to the numerals 1-3; namely *-jana* for human nouns and *-ta* for non-human nouns.

Nepali

dui-tā 'two-CLF' (Turnbull 1923: 51)

tin-əuta 'three-CLF' (Riccardi 2003: 559)

Chantyal

duy-ta 'two-CLF'

tin-ta 'three-CLF' (Noonan 2003:321)

Regarding the use of plural markers, Chantyal is not obligatorily marked for number (Noonan 2003: 318). Because the Chantyal language is under the influence of Nepali, it is quite possible that {N,NUM,NSG} might have developed due to contact with some Indic languages, such as Nepali which possesses the non-singular marker (Michael Noonan, personal communication). However, how the non-singular marker was borrowed is far from clear.

To summarise, like Malto, Chantyal exemplifies the spread of the numeral classifier systems via the borrowing of numeral systems (cf. Japanese in §8.3.1.2). Numeral borrowing seems to be a common factor which leads to the adoption of numeral classifier systems. As for cases like Chantyal, the adoption of a numeral classifier system is more easily conceivable (than in Malto/Japanese) because the numeral classifiers are bound to the numerals as a single word.

9.5 {N,NUM,OBL,SG}

The pattern of {N,NUM,OBL,SG} is another complex CNNC_{NSG} pattern. There are a couple of sources for the pattern of {N,NUM,OBL,SG}, namely {N,NUM,NSG} (§9.5.1) and {N,NUM} (§9.5.2).

9.5.1 {N,NUM,OBL,SG} < {N,NUM,NSG}

One possible explanation for the appearance of {N,NUM,OBL,SG} in a language is that it may have developed from {N,NUM,NSG}. Russian provides helpful evidence for this pattern. In Russian, {N,NUM,OBL,SG} (where the Obl is a genitive case) is used when the numerals are 2-4. Below is an example of this use with the numeral *two*.

(9.24) *Modern Russian (Neidle 1988: 95)*
dva *dnja*
two *day.GEN.SG*
'two days'

This pattern had developed from {N,NUM,NSG} in Old Russian (1100-1500 AD). According to Matthews (1960: 197-8) and Vlasto (1986: 232), the change occurred with the numeral *two* first and then extended to the numerals *three* and *four*.

In Old Russian, there were three grammatical numbers, namely singular, dual, and plural. The dual had been used in conjunction with the numeral *two*, as in (9.25a). At the beginning of the 13th century, the dual form, except in the nominative and accusative, was gradually lost and replaced by the plural, as shown in (9.25b). The nouns in the dual nominative and accusative forms were identical in general with the *genitive singular* form and hence the dual was reinterpreted as *genitive singular* (as in (9.25c)), although the nominative/accusative and genitive singular forms show different stress placement. Later, since the Muscovite period (14th-17th century), this process gradually extended to the numeral 'three' and the numeral 'four' which were previously followed by the plural.

(9.25) *Old Russian (Matthews 1960)*

a. *za* *měx̃* *dvě* *iogatě*
 for *fur* *two.DU* *pence.DU*
 ‘two pence for the fur’ (p.197)

b. *moi* *dva* *žerebʹja*
 1SG.POSS *two* *colt.PL*
 ‘my two colts’ (p. 198)

c. *dvě* *ženy*
 two *woman.GEN.SG*
 ‘two women’ (p. 198)

9.5.2 {N,NUM,OBL,SG} < {N,NUM}

This pattern of development indicates that {N,NUM,OBL,SG} may arise from {N,NUM}. This pathway is observed in Finnish, a Uralic language. Finnish uses the pattern of {N,NUM,OBL,SG} (where OBL is a partitive case) for CNNC_{NSG} as shown below.

(9.26) *Finnish (Nelson 1998:46)*

yhdeksän *omena-a*
nine *apple-PRTV*
‘nine apples’

The key question is what the pattern prior to {N,NUM,OBL,SG} was. Due to the shallow time depth of written documents in Finnish, there is no written evidence indicating the previous pattern of CNNC_{NSG} in the language. It is possible that the pattern of {N,NUM,OBL,SG} has existed since Proto-Finnic,³³ but due to the lack of written records we are not able to indicate the previous pattern of CNNC_{NSG} at the level of the language or even genus. So we may have to look at the level of the family.

³³The pattern of {N,NUM,OBL,SG} might have also been present in sister languages to Modern Finnish like Estonian (Virve-Anneli Vihman, personal communication).

Finnish is affiliated to the Uralic family. The existence of the Proto-Uralic family and the relationships among its main sub-branches are a matter of discussion (see Marcantonio 2002 and Bakro-Nagy 2005, for instance). However, the discussion in this section assumes that the Proto-Uralic family exists, and that Finnish is affiliated to the Uralic family. According to Koptjevskaja-Tamm and Wälchli (2001: 699-670), the original pattern of CNNC_{NSG} in Proto-Uralic has been under debate. This is because {N,NUM} is always present in languages which have had contact with Turkic languages, where {N,NUM} is dominant. Also, in some Uralic languages like Mordvin, spoken in Mordovia, Russia, nouns may be marked for plural after the numerals 2-10 (albeit infrequently). However, on the basis of data in a reference work *The Uralic Languages* (Abondolo 1998),³⁴ it seems that the Uralic languages often employ {N,NUM} as the primary means of forming CNNC_{NSG}. This is illustrated by Selkup, which belongs to the Samoyedic genus, spoken in Russia, and Hungarian, which belongs to the Ugric genus. Below are examples from the two languages.

(9.27) *Selkup* (Helinski 1998 575)

š'itti qum
two person
'two persons'

(9.28) *Hungarian* (Kenesei, Vago and Fenyvesi 1998: 229)

három fiú
three boy
'three boys'

The other two patterns, namely {N,NUM,OBL,SG} and {N,NUM,NSG}, are not productive. {N,NUM,OBL,SG} is limited to the languages around the Baltic sea. {N,NUM,NSG} is mentioned only in Nganasan (Samoyedic; Russia), but the use of the plural marker seems optional (Abondolo 1998).

³⁴ This reference describes the grammars of 18 languages representing various branches.

Therefore, in terms of productivity, the pattern of {N,NUM} in modern Uralic languages is the most likely the original pattern in Proto-Uralic because it is the pattern normally present in the family. However, further comparative research on the pattern of CNNC_{NSG} in all Uralic languages is required to confirm this hypothesis.

As for {N,NUM,OBL,SG}, which is used in Finnish and Estonian, this pattern is unlikely to be inherited from Proto-Uralic because the pattern is found only around the Baltic sea where the use of genitive case in CNNCs is commonly found, as is typical in Slavic & Baltic languages. Koptjevskaja-Tamm and Wälchli (2001: 698-701) believe that the use of partitive case in CNNC_{NSG} in Finnish can be attributed to contact with Baltic languages (e.g. Lithuanian) and Slavic languages (e.g. Russian). Koptjevskaja-Tamm and Wälchli also point out that the use of the partitive case in CNNC_{NSG} is similar to the so-called genitive case of Russian. That is, Slavic and Finnic numerals greater than *one* behave nominally in positions which require nominative or accusative case (i.e. the numeral is marked for the nominative or accusative case in the same way as nouns are, see example (9.29a)) and adjectivally in other positions (i.e. the case of the numeral agrees with nouns in the same way as adjectives do, see example (9.29b)).

(9.29) *Russian (Koptjevskaja-Tamm and Wälchli 2001: 698).*

a. *ja* *vižu* *[pjat'* *stakan-ov]*
 1SG.NOM *see.PRES.1SG* *five.NOM/ACC* *glass-GEN.PL*
 'I see five glasses.'

b. *ja* *prišla* *s* *[pjat'-ju stakan-ami]*
 1SG.NOM *come.PST.F.SG* *with* *five-INS glass-INS.PL*
 'I come with five glasses.'

This similarity suggests that the use of partitive case in Finnic is influenced by Slavic although it is not clear how the system is influenced. However, Larsson (2001: 244-247) argues on the basis of loanwords and certain syntactic similarities that the Finnic languages are in fact heavily influenced by Baltic languages, especially Lithuanian, since perhaps 3000-4000 years ago. The correspondence between genitive

case in Baltic and partitive case in Finnic is quite visible. This is evidenced by the examples from Finnish and Lithuanian below.

(9.30) *Finnish (Larsson 2001: 245)*

<i>juon</i>	<i>vettä</i>
<i>drink.1SG.PRES</i>	<i>water.PRTV.SG</i>
<i>'I drink water.'</i>	

(9.31) *Lithuanian (Larsson 2001: 245)*

<i>geriu</i>	<i>vandens</i>
<i>drink.1SG.PRES</i>	<i>water.GEN.SG</i>
<i>'I drink water.'</i>	

Turning to $CNNC_{NSG}$, the use of genitive case in Lithuanian corresponds to the use of partitive case in Finnish. The noun after numerals above *one* is assigned genitive (plural) case in Lithuanian and partitive (singular) case in Finnish as shown below.

(9.32) *Lithuanian (Ambrazas 1997: 587)*

<i>vienuolika</i>	<i>vaikų</i>
<i>eleven</i>	<i>child.GEN.PL</i>
<i>'eleven children'</i>	

(9.33) *Finnish (Nelson 1998:46)*

<i>yhdeksän</i>	<i>omena-a</i>
<i>nine</i>	<i>apple-PRTV.SG</i>
<i>'nine apples'</i>	

If we believe that the $CNNC_{NSG}$ in Finnic (here Finnish) is influenced by the Baltic languages, a question may arise why the nouns in Finnish appear in the singular, unlike Lithuanian where they appear in the plural. There is no clear answer to this question. However, this might be because it is a remnant of the feature of Proto-Uralic which assigns the singular to nouns after numerals.

9.6 {N,NUM,OBL,NSG}

This pattern consists of four constituents—namely, a noun, a numeral, an oblique case (including prepositions such as *of*), and a non-singular marker. There are two patterns of development for this complex construction. One is {N,NSG} and the other is {N,NUM}. The former is illustrated by Modern Standard Arabic (Afro-Asiatic; Middle East and North Africa) (§9.6.1), the latter by Welsh (Indo-European; Wales), discussed in §9.6.2.

9.6.1 {N,NUM,OBL,NSG} < {N,NSG}

This pattern of development denotes that {N,NUM,OBL,NSG} develops from {N,NSG} (where NSG is a plural marker). We provide examples here from Modern Standard Arabic in this regard.” Of the several CNNC_{NSG} patterns in this language (namely {N,DU}, {N,NUM,NSG}, {N,NUM,OBL,SG}, and {N,NUM,OBL,NSG}), the one which concerns us here is the pattern of {N,NUM,OBL,NSG} (where the oblique is the genitive case), as shown in the example below.

(9.34) *Modern Standard Arabic (Holes 1995)*

<i>ʔarbáat-u</i>	<i>ʔawla:d-in</i>
<i>four.F-NOM</i>	<i>boy.PL-GEN</i>
<i>‘four boys’ (p.173)</i>	

It is argued that {N,NUM,OBL,NSG} does not in fact develop from any already existing CNNC_{NSG}. Based on the universal tendency for oblique case to be used for high numerals, proposed in Corbett (1978) (cf. §5.3.4), it is conjectured that the source for this pattern is {N,NSG}. That is to say, the noun was first obligatorily accompanied by the non-singular (cf. §5.13, 4th degree), and then once the new high numerals emerged in the language, these numerals were therefore used with the {N,NSG} construction. The oblique case is required because these high numerals are treated as nouns (not as adjectives like low numerals), and the language does not allow double nominatives.

Therefore, {N,NUM,OBL,NSG} does not derive from any structural types of CNNCs (except in the case of Welsh, discussed below).

In Modern Standard Arabic, numerals vary in word classes and in their syntactic behaviour. The numerals 1 and 2 behave adjectivally, as they are placed in the same syntactic slot as the modifier. That is to say, the modifier follows the head noun. See the examples (9.35) and (9.36). The example (9.35a-b) illustrates that the word *new* is an adjectival modifier and precedes the noun only.

(9.35) *Modern Standard Arabic* (Shlonsky 2004: 1470)

a *Volvo* *xadaš-a*
 Volvo *new-F*
 'a new Volvo'

b **xadaš-a* *Volvo*
 new-F *Volvo*
 'a new Volvo'

(9.36) *Modern Standard Arabic* (Holes 1995: 173)

a. *kita:b* *wa:ħid*
 book *one*
 'one book'

b. *kita:b-a:ni* *tna:ni*
 book-DU *two.M*
 'two books'

The other numerals (i.e. those greater than 2) behave nominally; they are treated as the head of the noun phrase—that is, they precede the modifier. Consider the example (9.34) in conjunction with (9.35) above.

Another piece of evidence which suggests that numerals greater than 2 behave nominally in Modern Standard Arabic is that they can receive case (such as nominative case in the subject position), whereas the numerals 1 and 2, being adjectival, are not marked for case.

The use of {N,NUM,OBL,NSG} for numerals greater than 2 also existed in Classical Arabic (8th-10th centuries) as shown below. It can also be noted that non-numeral quantifiers require quantified nouns in the genitive, as shown in (9.37a-b).

(9.37) *Classical Arabic* (Benmamoun, 2003: 761)

- a. *xamsat-u* *kutub-in*
 five-NOM *book.PL-GEN*
 ‘five books’
- b. *kull-u* *t-tullaab-i*
 all-NOM *DEF-student.PL-GEN*
 ‘All the students’

Overall, we can see that numerals vary in their syntactic behaviour. Following the universal tendency proposed in Corbett (1978), when numerals greater than 2 are combined with quantified nouns, the genitive case is then required to avoid double nominatives. The interesting issue is why the numerals from 3 onwards are in the head position, whereas the numerals 1 and 2 are in the modifier position. As the two groups of numerals both function as quantifiers, they should be expected to be in the same syntactic slot. The fact that they occur in different positions suggests that they might have existed at different times; otherwise the numerals 2 and 3, which are not so different from each other, would be in the same slot. It is perhaps coincidental that in Sumerian, an extinct language of the ancient Middle East (4000 BC), the numeral 3 and the plural marker share the same form: the forms for the numeral 3 are *eš* and *peš* and the plural marker is *eš* (Schmandt-Besserat 1996: 113). Sumerian numeration is also based on a ternary system (i.e. a three-count system), for example, *pešbala* [*three-passed*] ‘four’. This suggests that the most basic numerals are 1 and 2 in the language, whereas the numerals 3 and beyond are treated as ‘lots’ or ‘beyond counting’.

9.6.2 {N,NUM,OBL,NSG} < {N,NUM}

This pattern of development denotes that {N,NUM,OBL,NSG} derives from {N,NUM}. Welsh is an example illustrating this pattern. In Welsh, there are two CNNC_{NSG} constructions, namely {N,NUM}, as in (9.38a) and {N,NUM,OBL,NSG}, as in (9.38b). Generally, the former is used with low numbers and the latter is used with higher numbers. It is hard to draw the dividing line between the two constructions though. The number 10 is sometimes suggested. However, at present, the construction {N,NUM,OBL,NSG} is also used even with low numbers like 2 or 3 (King 1993:111).

(9.38) *Welsh*

a. *tri dyn*
three man
 'three men' (King 1993:111)

b. *tri o ddynion*
three of man.PL
 'three men' (Watkins 2003: 332)

Following the universal tendency that languages where cardinal numerals vary in syntactic behaviour (i.e. some numerals are adjectives and some are nouns), the pattern of {N,NUM,OBL,NSG} tends to be used with high numbers, not low numbers. Accordingly, the use of {N,NUM,OBL,NSG} with low number in Welsh is presumably innovative. It is noted in Hurford (2003: 12) that:

What makes Welsh unusual is the extension of this kind of structure [{N,NUM,OBL,NSG}] to indefinites and to low-valued numerals. Such constructions with numerals, prepositions and indefinite nouns, semantically equivalent to ordinary attributive numeral-noun structures, are relatively rare, in Europe at least. Where such prepositional constructions do occur, it tends to be with the most high-valued, nounier numerals.

The motivation for the extension of {N,NUM,OBL,NSG} to low-valued numerals is not clear. This phenomenon can perhaps be seen as a matter of *free variation*. Even the same speaker may tend to use the two constructions alternately. For example, as noted in King (1993: 111), for ‘two children’ they may prefer to say *dau o blant* [two-of/from-child.PL] rather than *dau blentyn* [two-child]. In any case, the extension may involve syntactic analogy. By analogy with high-valued numerals, the structure of {N,NUM,OBL,PL} may be extended to the low-valued numerals. The preference for the structure plays an important role in driving this syntactic analogy.

10 Evolutionary Trajectories of CNNCs

In Chapter 6 (CNNCs Across Languages), we observed a variety of structural patterns of CNNCs in the world's languages. Then, in Chapter 8 (Historical Origins of CNNC_{SG}) and Chapter 9 (Historical Origins of CNNC_{NSG}), these structural patterns were further investigated in order to examine the structural patterns immediately preceding these contemporary patterns. According to the cross-linguistic comparative approach integrated with language history and theoretical diachronic frameworks, in this chapter, the various historical pathways of CNNCs are combined together to postulate larger subsystems in several layers. We then end up with *possible* general evolutionary trajectories of CNNC_{SG} and CNNC_{NSG} , as presented in §10.1 and §10.2 respectively. The synthesis and integration of the findings are based on the assumption that a variety of the patterns of CNNCs in the world's modern languages can reveal various stages of evolutionary development of CNNCs from the initial stages to more recent ones. To emphasize, the reconstruction of the evolutionary scenarios of CNNCs proposed here is just one possibility and is hypothetical to some extent. The evolutionary trajectories of CNNCs are followed by a discussion of the contributing factors that influence the complexity and diversity of CNNCs (§10.3). Finally, the nature of the evolution of CNNCs is discussed (§10.4). Note that since this chapter is treated as giving an overall picture of preceding chapters, some issues and examples already mentioned in those chapters may be repeated here for convenience.

10.1 Evolutionary trajectory of CNNC_{SG}

This section proposes a possible evolutionary scenario of CNNC_{SG} derived from the findings on historical paths of each structural type of CNNC_{SG} from Chapter 8. The evolutionary scenario of CNNC_{SG} can be summarized as in Figure 10.1 below. The languages representing the historical paths are shown in square brackets.

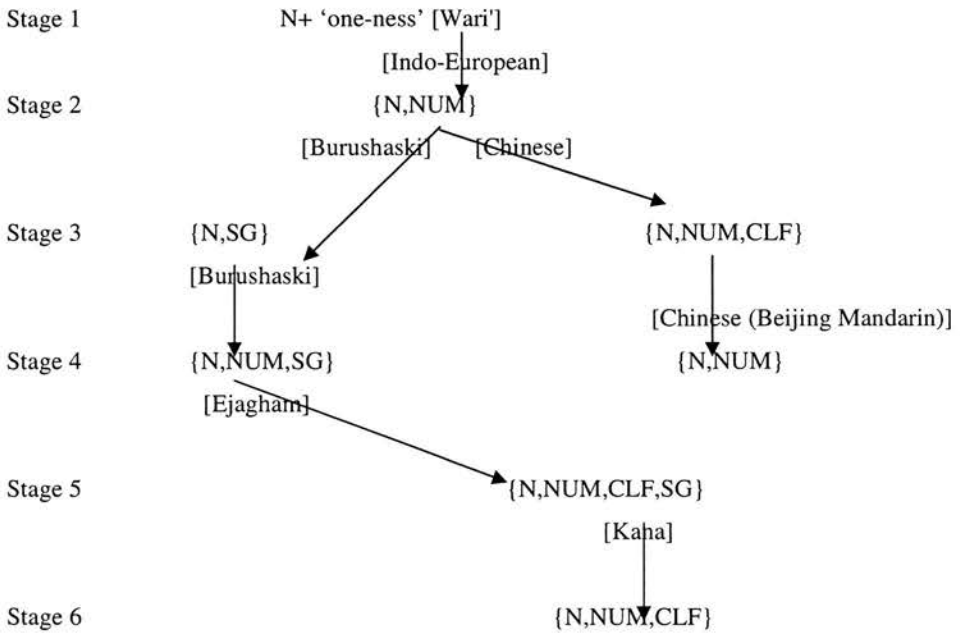


Fig. 10.1 A Possible Evolutionary Trajectory of CNNC_{SG}

The schema presented in Fig. 10.1 describes a possible evolutionary scenario of CNNC_{SG}. At the first stage, before the emergence of the numeral 1 proper (i.e. before the rise of the CNNC_{SG}), the words denoting 'one-ness' such as 'alone' and the like, as in Wari' (Everett and Kern 1997: 347-348), might have been used to express the one-ness of things. This includes the use of the body-part counting system which is common in Papua New Guinea. This stage where the numeral 1 proper was not yet present may be referred to as the *embryonic stage* of CNNC_{SG}. Languages which do not develop beyond this stage do exist but are extremely rare. These languages provide very helpful

evidence for the first stage. From this stage onwards, the $CNNC_{SG}$ has split into several types over time.

From Stage 1 to Stage 2, the use of the words conveying the sense of ‘one-ness’ was extended to designate the numeral 1 through a process of semantic extension. Since then, {N,NUM} which is the most basic pattern of $CNNC_{SG}$ became apparent in languages. The transition as such is attested in the histories of living languages. For example, in the Indo-European languages, the root of the reconstructed form of the numeral 1 (*oi-*) etymologically means ‘alone’ (Burrow 2001: 258). In the Papua New Guinean languages, such as Haruai, the numeral *aglŋ* ‘one’ is apparently cognate with the word meaning ‘little finger’ (Comrie 1999: 91). It is possible that the pattern of {N,NUM} in a number of languages might have arisen in the same way as happened in the Indo-European languages and Haruai³⁵. That is, the numeral 1 was derived from the lexical word meaning ‘one-ness’ or from a body-part term, typically a word meaning ‘finger’. From the fact that the pattern of {N,NUM} is the most common type in the world’s languages (cf. §6.1), it is reasonable to infer from this fact that the development of $CNNC_{SG}$ in most languages stops at the second stage. However, note that some instances of the pattern of {N,NUM} may have developed from other patterns. For example, in Beijing Mandarin Chinese, the pattern of {N,NUM} developed from {N,NUM,CLF}. This will be further discussed in the fourth stage.

From Stage 2 to Stage 3, in some languages, {N,NUM} keeps developing towards two different constructions, namely {N,SG} and {N,NUM,CLF}. The former pathway is attested in Burushaski (isolate; Pakistan) and the latter in Beijing Mandarin Chinese. In Burushaski, the pattern of {N,NUM} may develop into {N,SG} via the process of grammaticalization when the numeral 1 is used as a singulative marker instead. For example, Burushaski *gərk* ‘peas’ becomes *gərkən* ‘a pea’. The singulative suffix ‘-en’ is derived from the numeral *hən* ‘one’ (Lorimer 1935: 191). Thus, the

³⁵ It is noted in Comrie (1999: 81) that the system might have been borrowed from Kobon, a neighbouring New Guinean language from which a sizeable vocabulary in Haruai is evidently borrowed. Although Haruai borrowed the body-part system from Kobon, however, this does not affect our conclusion. The crucial point of interest is that the body-part system reflects the pre-stage of $CNNC_{SG}$, no matter in what language the system was originally used.

singulative marker functions as an individualizer. That is, the singulative marker is used to individualize the collective noun 'peas'. This phenomenon can happen in other languages where some nouns in the language behave as collective nouns. This quantifying function of the singulative marker leads to the rise of the {N,SG} construction in the language.

Turning to Beijing Mandarin, the development of {N,NUM,CLF} from {N,NUM} took place when the nouns denoting a discrete spatial entity (for example, the noun meaning 'ox') were reinterpreted in the same way as the nouns denoting a group of entities, for example, the nouns meaning 'shells' and 'jades' which were culturally perceived as a group and a pair, respectively. These kinds of nouns are generally referred to as collective nouns, like English *cattle*. When the nouns denoting a group of entities are quantified—that is, in combination with a numeral or a quantifier, the numeral classifier is then required to make those nouns enumerable.

In addition, language contact may lead to the rise of {N,NUM,CLF} in some languages. The typical situation is when a language borrows numerals from another language. This can be illustrated with Japanese. In Japanese, although it is not clear whether this language had its own numeral classifier system, the Japanese numeral classifier system developed evidently when the Chinese numerals were borrowed into the language. This can be seen by the fact that numeral classifiers are often used with the Sino-Japanese numerals and are hardly used with the traditional Japanese numerals (Downing 1996: 35-51).

Later, from Stage 3 to Stage 4, at the side of the diagram (Fig. 10.1) with {N,SG} when the numeral 1 has changed its function to a singulative marker, it is then no longer the numeral 1. When the speaker wants to apply the number 'one' to the referent, the numeral *han* 'one' must be used. This is why the numeral 1 co-occurs with the singulative marker despite the fact that the forms of the numeral 1 and the singulative marker are not so different. Once again, this pathway is illustrated by Burushaski.

Turning to {N,NUM,CLF}, {N,NUM,CLF} may develop into {N,NUM}. This pattern is attested in Beijing Mandarin. In this language, the numeral classifiers may be fused with numerals via the integral process of grammaticalization and lexicalization,

giving rise to the revival of the pattern of {N,NUM}. This phenomenon can happen when the numeral and the numeral classifier have a high frequency in terms of their co-existence. When the whole system of numeral classifiers is fused with the numerals, the numeral classifiers as free morphemes will no longer exist in the language. Therefore, the pattern of {N,NUM} as observed in the languages of today is not necessarily the original pattern, rather the pattern may be a new one resulting from a recent development.

From Stage 4 to Stage 5, the pattern of {N,NUM,SG} may develop towards {N,NUM,CLF,SG}. This pathway is attested in Ejagham. In Ejagham, when there are some nouns relevant to trees or plants which can be counted in various ways, numeral classifiers will be required. Since the noun class/number markers still function, however, this leads to the pattern of {N,NUM,CLF,SG}.

Finally, from Stage 5 to Stage 6, the pattern of {N,NUM,CLF,SG} can develop further to {N,NUM,CLF} by dropping the noun classes which include number markers, as happened in Kana.

10.2 Evolutionary trajectories of $CNNC_{NSG}$

This section proposes a possible evolutionary scenario of $CNNC_{NSG}$ derived from the findings on historical paths of each structural type of $CNNC_{NSG}$ from Chapter 9. The evolutionary scenario of $CNNC_{NSG}$ can be summarized as in Figure 10.2 below. The languages representing the historical paths are shown in square brackets.

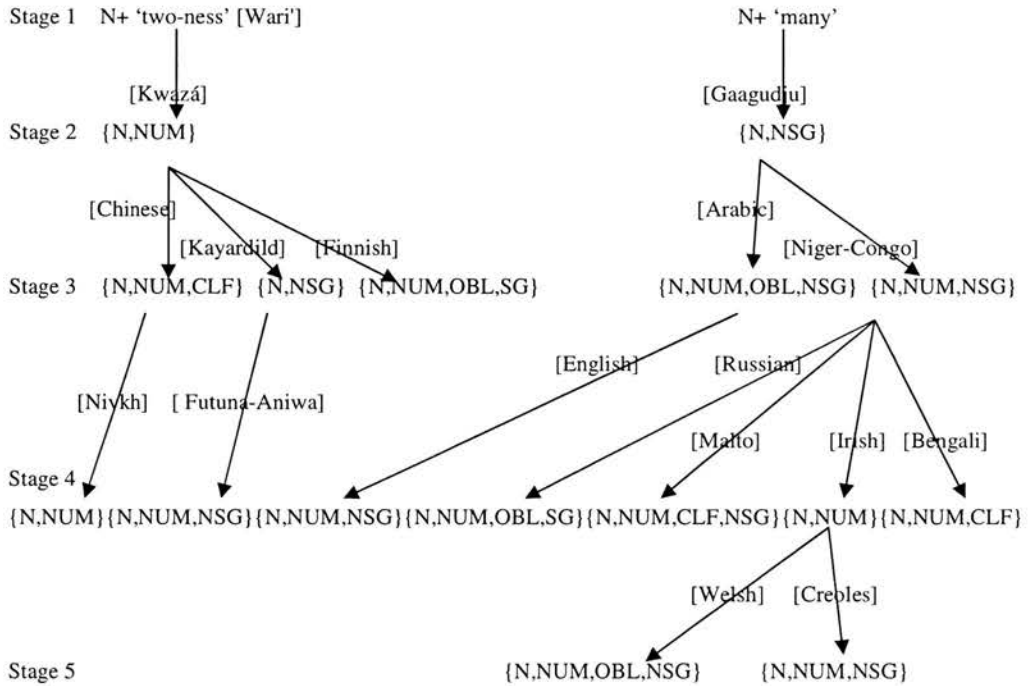


Fig. 10.2 A Possible Evolutionary Trajectory of $CNNC_{NSG}$

The diagram shows that at the initial stage, before the emergence of numerals greater than 1, the word denoting 'two-ness', such as Wari (Chapacura-Wanhan; Brazil) *tucu caracan* 'face each other' might have been used to express the two-ness of things. This stage where the word with a numerical interpretation is used to express a non-singular concept may be referred to as the *embryonic stage* of $CNNC_{NSG}$ because the stage is treated as a foundation for the next stage, namely $\{N, NUM\}$. Consider the right

trajectory; the words denoting non-singularity may be used, for example, ‘many’, ‘all’, and ‘they’.

From Stage 1 to Stage 2, the words denoting ‘two-ness’ can develop into the numeral 2. The phenomenon whereby the numeral 2 develops from the lexical word denoting ‘two-ness’ is evidenced by the etymological study of numerals. For example, in the Eskimo-Aleut languages, etymologically, the numeral 2 (e.g. Alaska *malruk* ‘two’) presumably developed from the word meaning ‘follow’ (Bonnerjea 1978: 53). In Kwazà (Kwaza; Brazil; Voort 2004: 214), the word meaning ‘two’ share the same form of the word meaning ‘company’. Once the lexical word denoting ‘two-ness’ develops into the numeral ‘two’ proper and is used with the quantified noun, then the pattern of {N,NUM} comes into being in the language. It is possible that {N,NUM} in most languages may arise in the same way as happened in Wari’ or Kwazà. In some languages, the CNNC_{NSG} does not develop beyond Stage 2 (i.e. {N,NUM}), whereas in other languages the development continues as illustrated below.

Regarding other numerals, according to etymological evidence, they arose through two processes. The first process is *numeralization*. The meaning of the numeral 3, for example, in Indo-European languages (e.g. English ‘three’, French ‘trois’) is extended from the word **ter* meaning ‘beyond’ (Luján Martínez 1999: 207). The second process is a combination of the already existing numerals, i.e. *one plus two*; and the higher numerals are built on earlier numerals, for example, Baruya (Trans-New Guinea; Papua New Guinea) *dawaai-da* [one-two] ‘three’ (Phythian 2007: 3).

At the right-hand trajectory, the words denoting plurality, such as ‘many’, ‘all’ and ‘they’ developed into plural markers through the process of grammaticalization. When the grammatical number is obligatorily attached to the nouns in the context of plurality, it can be said that the pattern of {N,NSG} (e.g. English *dog-s*) has arisen. However, as already mentioned in §4.3, there might have been other means of number distinction, such as changes within the noun stem, plural tone, plural words, and reduplication (Dryer 2005a). For example, in Gaagudju (Australian, Northern Territory), the morpheme *mana*, which is the *masculine unit augmented clitic* (MUA), is used to express plurality, e.g. *anmarrabaalbu=mana* [oldman=MUA] ‘old men’ (Harvey 2002).

In some languages, these means of number distinction may become obligatory in the non-singular context over time. Overall, the stage of {N,NSG} is the embryonic stage of the pattern of {N,NUM,NSG} in later periods.

From Stage 2 to Stage 3, {N,NUM} and {N,NSG} may evolve in various directions. As for {N,NUM}, the first direction is that {N,NUM} develops towards {N,NUM,CLF} as happened in Chinese. The process of change has already been described in CNNC_{SG}. The reader is referred back to §10.1, Stage 3-Stage 4, in this regard. The second direction is the development towards {N,NSG} (NSG here is a dual marker or a trial marker; but not a plural marker), when the numerals 2 and 3 are grammaticalized into the dual and trial markers respectively. This pathway can be observed in the Australian and Oceanic languages. For example, in Kayardild (Australian, Queensland), the dual marker *jiyarrng* is grammaticalized from the numeral *kiyarrng* 'two' (Evans 1995: 184). The third direction is the development towards {N,NUM,OBL,SG}. This pathway can be found in the Finnic languages, such as Finnish. The development can be regarded as a *language-specific idiosyncrasy*. In this case, the pattern arises because the language had been in contact with the languages where the oblique marker is used in CNNC_{NSG}.

At the right-hand trajectory, the pattern of {N,NSG} (where NSG is a plural marker only) developed into the pattern of {N,NUM,NSG} after the numerals proper were introduced in the language. This pathway is hypothetical to some extent. Speculating somewhat (due to the lack of concrete evidence), we can conjecture that the non-singular marker might have been obligatorily attached to the noun in the context of non-singularity. In English, the plural marker '-s' is always required in the context of non-singularity even though the numerals are not present. This reflects the possibility that the plural nouns (the nouns in plural forms) or {N,NSG} might have existed before the arrival of numerals.

Turning to the pattern of {N,NUM,OBL,NSG}, this pattern emerged when the high numerals, especially high round numerals were introduced in a language where the higher numerals tended to be nouns and the two nouns were not allowed to occur next to each other. This phenomenon can be observed in Arabic and some Indo-European

languages, for instance, where the high round numerals require an oblique marker when combined with nouns.

From Stage 3 to Stage 4, the pattern of {N,NUM,OBL,SG} does not develop any further. However, some structural patterns still keep developing into other patterns. The pattern of {N,NUM,CLF} may develop into {N,NUM}(or exactly {N,NUM.CLF}), when the numeral classifiers are completely fused with the numerals in the language. This structural type is illustrated by Nivkh (isolate; Siberia, Russia Gruzdeva 1998: 24) where numerals are totally fused with numeral classifiers, giving rise to several sets of numerals, for example, the numeral *men* 'two' (for counting people) and the numeral *mor* 'two' (for counting animals). The phoneme /m/ seems likely to be the root of the numeral 2.

The pattern of {N,NSG} where the NSG is a dual marker or trial marker developed into {N,NUM,NSG} when the dual or trial markers assumed some other grammatical functions (e.g. as an article) and the meaning concerning the quantifying function (i.e. expressing the number *two* or *three*) was bleached. So, the numerals 2 and 3 would have come into use again. Futuna-Aniwa (Austronesian; Vanuatu) is an example of such a language, where the numeral 2 has become re-used once the dual marker became grammaticalized into the dual article (Dryer 1989: 869).

At the right-hand trajectory, {N,NUM,OBL,NSG} developed into {N,NUM,NSG} when the high round numerals came to be treated more adjectivally, and hence the oblique was no longer required. This is attested in at least some Indo-European languages, such as English. Regarding the pattern of {N,NUM,NSG} which developed from {N,NSG}, it developed into various patterns. Firstly, it developed into {N,NUM,OBL,SG}. This pathway is attested only in Russian. The change happened first after the dual form was reinterpreted as the genitive singular. Later, the pattern of {N,NUM,OBL,SG} extended its use from the numeral 2 to the numerals 3 and 4 (Matthews 1960: 197-8 and Vlasto 1986: 232). This is another pattern of development regarded as a *language-specific idiosyncrasy*.

The pattern of {N,NUM,NSG} can develop into {N,NUM,CLF,NSG}. This pathway is illustrated by Malto, a Dravidian language (cf. Steever 1998: 372). The pattern of {N,NUM,CLF,NSG} developed in the language after the language adopted the Indo-Aryan numeral classifiers accompanying the Indo-Aryan numerals.

The pattern of {N,NUM,NSG} can also develop into {N,NUM} after the non-singular marker was dropped. This is attested in Irish when the case marker which simultaneously denoted grammatical number was dropped for phonological reasons. Finally, the pattern of {N,NUM,NSG} may develop into {N,NUM,CLF}. This path is evidenced in Bengali (Indo-European; India), a language descended from Sanskrit which has the pattern of {N,NUM,NSG}.

Finally, at the fifth stage, {N,NUM} may develop further into {N,NUM,NSG} as observed in pidgins and creoles, perhaps, for a reason of standardization. The pattern of {N,NUM} also develops into {N,NUM,OBL,NSG}, perhaps by syntactic analogy, as evidenced in Welsh. The latter path is rather unusual, since it is found nowhere else. Both developmental paths reflect the most recent change in $CNNC_{NSG}$.

10.3 Motivations for the origins of CNNCs

The preceding sections describe how CNNCs evolved. In this section, to answer the question of why CNNCs developed in some ways, the motivations for the rise of various structural patterns of CNNCs are discussed. The changes caused by these factors can be divided into three categories, namely quantifying function-motivated change (§10.3.1); non-quantifying function-motivated change (§10.3.2); and a mixture of both (§10.3.3).

10.3.1 *Quantifying function-motivated change*

Under this heading, the changes which might plausibly be attributed to pressures to express quantification or numerical meanings successfully are considered. They include the rise of numerals and the growth of collective nouns and general nouns. This kind of change is referred to as *quantifying function-motivated change*.

10.3.1.1 Rise of numerals

The rise of numerals in a language, either derived from the existing words in the language through numeralization or borrowing from another language, can lead to the emergence and development of CNNCs. Since the rise of numerals is relevant to quantification, it is regarded as a factor contributing to a *quantifying function-motivated change*. The following are the structural patterns motivated by the rise of numerals.

(a) {N,NUM}

This pattern is derived from the noun phrase construction consisting of a noun and a word with a numerical meaning. When these words were *numeralized* via semantic extension, the numerals then arose in the language. Consequently, the pattern of {N,NUM} emerged. Therefore, the emergence of this pattern is attributable to the rise of numerals.

(b) {N,NUM,NSG}

This pattern presumably developed from the construction consisting of a noun and a non-singular marker (i.e. the pattern of {N,NSG}, such as English *dogs*). The pattern of {N,NSG} did not convey a clearly exact number. In other words, the

construction of {N,NSG} emerged first when the language did not have numerals proper at all. Then, once the numerals arose in the language, they were combined with {N,NSG} later to fulfil a functional need for exact quantification. So, the pattern of {N,NUM,NSG} is another pattern which emerges in a language when numerals appear. It is therefore regarded as an instance of the quantifying function-motivated change.

(c) {N,NUM,OBL,NSG}

The pattern of {N,NUM,OBL,NSG} where the oblique case is a genitive/partitive case/preposition (e.g. Old Norse *fimm hundruð manna* [five-hundred.NSG-man.GEN.NSG] ‘five hundred men’) is presumably derived from the construction consisting of a noun and a non-singular number (i.e. the pattern of {N,NSG} here is *manna* ‘man.PL’). In this case, the pattern of {N,NUM,OBL,NSG} arises when the high round numerals, such as ten or a hundred, are introduced in the language and these high round numerals are treated as nouns. Therefore, the rise of {N,NUM,OBL,NSG} is attributed to the rise of high round numerals.

(d) {N,NUM,CLF}

The pattern of {N,NUM,CLF} arose in some languages when those languages borrowed numerals from other languages. The numeral classifiers may come along with these numerals, as has happened in the Japanese and Munda languages. In these two languages, the numeral classifiers are always used with the borrowed numerals. Hence, in this case, the rise of the pattern of {N,NUM,CLF} is due to borrowing of numerals and the change can be regarded as a quantifying function-motivated change because the change deals with the rise of (new) numerals in the languages.

(e) {N,NUM,CLF,NSG}

This pattern is derived from the pattern of {N,NUM,NSG}. The change is attested in Chantyal (Sino-Tibetan; Nepal) where the numeral classifiers have come into use in the language because the numeral classifiers are attached with the borrowed numerals. This results in the structure of {N,NUM,CLF,NSG}, for example, *tin-ta*

'three-CLF' (Noonan 2003:321). The numeral plus classifier *tin-ta* is borrowed from Nepali *tin-əuta* 'three-CLF' (Riccardi 2003: 559). Therefore, the rise of the pattern of {N,NUM,CLF,NSG} is also due to borrowing of numerals. Like {N,NUM,CLF}, the rise of the pattern of {N,NUM,CLF,NSG} is attributed to the arrival of new numerals in the languages and so the change can be regarded as a quantifying function-motivated change.

10.3.1.2 Presence of collective nouns and general nouns

In the languages where collective and general nouns are found, the linguistic items functioning as what are generally referred to as *unitizers* or *individualizers* are always present as well. The unitizers may be a singular or singulative marker grammaticalized from the numeral 1. The unitizers can be a numeral classifier grammaticalized from a noun. In any case, the growth of collective nouns or general nouns in languages leads to the rise of new types of CNNCs. This factor is regarded as involving quantification, as the singularizer behaves as a unitizer, making the noun enumerable. The structural patterns motivated by this factor are:

(a) {N,NUM,SG}

This pattern developed from {N,SG}, which in turn developed from {N,NUM}, where Num is the numeral 1. For example, Burushaski *huyes* 'goat, sheep' can be singularized by the suffix *-an* (which is grammaticalized from the numeral *han* 'one', becoming *huyesan* 'a goat, sheep'). Then the word is combined with the numeral 1, becoming *han huyesan* [one-goat, sheep.SG] - 'one goat, sheep' (Lorimer 1935:48). The emergence of the pattern of {N,NUM,SG} is regarded as a result of a quantifying function-motivated change. This is because there is a need to count things in groups, so the singularizer is needed to make the enumeration possible.

(b) {N,NUM,CLF}

This pattern developed from the pattern of {N,NUM} as in Chinese. In Chinese, when the discrete nouns (i.e. nouns denoting referents with a discrete boundary) were changed to the collective-like nouns, then the numeral classifier became required to individualize nouns when the nouns were quantified. This is a possible origin of numeral classifiers or {N,NUM,CLF}. Hence, it can be said that the rise of {N,NUM,CLF} results from a quantifying function-motivated change.

10.3.1.3 Presence of nouns counted in more than one way

There is a group of nouns in languages which can be counted in more than one way. These nouns mostly denote *plants*. They can be counted as *fruits* or *trees*, for instance. For example, in Ejagham, the stem -*čókùd* ‘orange’ can be counted as a tree or as a fruit. For this reason, these nouns require a numeral classifier functioning as a unitizer. The presence of this kind of noun in languages can account for the existence of numeral classifiers or {N,NUM,CLF} in some languages. The only pattern motivated by this is the pattern of {N,NUM,CLF} which is derived from the pattern of {N,NUM,NSG}, as happened in Ejagham. The phenomenon is attributed to a quantifying function-motivated change because it involves counting.

10.3.1.4 Special meaning

In a language where the patterns of {N,DU} and {N,NUM} (where the numeral is ‘two’; see the examples (10.1a-b) below) are both used to convey the meaning of ‘two entities’, the meaning of the two patterns may be slightly different. In Ngiyambaa (Pama-Nyungan; New South Wales), {N,DU} conveys the meaning of *a set of two* (or a pair), whereas the pattern of {N,NUM} conveys the meaning of ‘two entities’ only (Donaldson 1980: 102).

(10.1) *Ngiyambaa* (Donaldson 1980: 102)

a. {N,NUM}

bulagar miri

two dog

'two dogs'

b. {N,DU}

miri-bula:

dog-DU

'a pair of dogs'

10.3.2 Non-quantifying function-motivated change

The non-quantifying function-motivated change refers to a change dealing in no way with counting, but purely with grammatical issues. The changes in CNNCs to be described as follows are hard to ascribe to quantifying function-motivated changes. Below is a list of language changes contributing to the change in CNNCs and the structural patterns of CNNCs motivated by these language changes.

10.3.2.1 Loss of cases and number

The loss of cases may trigger change in CNNCs. This is illustrated by the Goidelic languages (e.g. Irish and Scottish Gaelic). The Goidelic languages, following the traditional Indo-European pattern, presumably had the pattern of {N,NUM,NSG}. However, some cases originally marked on the unaccented final syllable were dropped and hence, the number marking was dropped accordingly (Acquaviva 2006:1866). The loss of cases represents change unrelated to counting.

Besides, in pidgins and creoles, the non-singular markers are mostly absent from CNNCs. This is because the inflections such as gender, tense and number of those Indo-European languages which are their parent languages are optional or are dropped. For that reason, the grammatical number may disappear not only in the context of CNNCs but also when used with the nouns mentioned in general as in (10.2 a-b).

(10.2) *Jamaican Creole*

(a) *di tu book*
DEF.ART two book
'the two books' (Bailey 1966: 30)

(b) *manggo swiit*
mango sweet
'mangoes are sweet' (Bailey 1966: 27)

In sum, the absence of the non-singular marker in the numeral context cannot be attributed to a quantifying function-related change, but to a simplified grammar which is a characteristic of pidgins and creoles.

10.3.2.2 Fusion of numeral classifier and numeral

A fusion of the numeral classifier and numeral may trigger the change in CNNCs. This can be illustrated by Beijing Mandarin. Beijing Mandarin has the pattern of {N,NUM,CLF}. However, due to the fusion of a numeral classifier and certain numerals, the numeral classifier is fused with the numeral, resulting in the pattern of {N,NUM}. The fusion of a numeral classifier and a numeral represents a factor of change in CNNCs which does not involve quantification, but rather represents sound change (i.e. assimilation). The only structural pattern motivated by this factor is {N,NUM} which derives from the pattern of {N,NUM,CLF}.

10.3.2.3 Change in grammatical categories of numerals

The pattern of {N,NUM,NSG} (e.g. Modern English *four hundred winters*) is derived from the pattern of {N,NUM,OBL,NSG} (e.g. Old English *four hundred of winters*). The change is illustrated using Old English. In Old English, the high round numerals such as 'hundred' are treated as nouns, but in Modern English, these numerals are treated as adjectives. So, the oblique case (genitive case here) is no longer required. The change in grammatical categories of the numerals is regarded as a non-quantifying function-related change because the grammatical category of high round number is

changed from a noun to an adjective. This factor is not regarded as dealing with quantifying function but rather with a grammatical matter.

10.3.2.4 Reanalysis of grammatical elements

The change in CNNCs may be attributable to a historical accident such as took place in Old Russian when some nouns in the dual nominative and accusative forms were reinterpreted as *genitive singular*. The phenomenon occurred because the dual forms looked identical in general to the *genitive singular* form. Obviously, no quantifying function-motivated change is involved in this change.

10.3.3 *Quantifying function-motivated change plus non-quantifying function-motivated change*

The *quantifying function-motivated change plus non-quantifying function-motivated change* refers to a change concerning counting together with grammatical matter. Below is a list of factors of this kind and the structural patterns of CNNCs motivated by these factors.

10.3.3.1 Loss of agreement

The loss of the agreement or concordial system may trigger the change in $CNNC_{NSG}$ as evidenced in Arabic. In Classical Arabic, the dual marker functions as a grammatical agreement category. Later, the concord was lost in the Arabic dialects. The dual was then marked on nouns only, not on other elements in the construction, making the dual at this stage indistinguishable from the numeral 'two'. When the numeral 'two' was present, the dual became redundant and was omitted. The use of a plural marker which, in its basic form, was used to mark plural nouns (more than 'two') became extended to the context with the numeral 'two'. The process resulted in the pattern of {N,NUM,NSG} (where the non-singular marker is a plural marker only, not trial or dual). This case involves a quantifying function-motivated change—that is, the dual is dropped in the context of the numeral 2. Also, the change involves a grammatical issue, namely the loss of concord or agreement in the language. The only structural pattern

motivated by these two factors together is {N,NUM,NSG} which developed from the pattern of {N,NSG} (where NSG is the dual marker).

10.3.3.2 Grammatical change in the number marker

When the number marker is changed by way of its function, this may trigger a change in some structural types of $CNNC_{NSG}$ where the number marker is present. For example, in Futuna-Aniwa (Austronesian; Vanuatu), the dual marker was used instead of the numeral 'two'. However, when the dual marker assumed the function of an article, the numeral 2 was required again. This is another case showing change involving both quantifying function-related change (i.e. when the numeral is re-used to clarify the number of the referent) and non-quantifying function-motivated change (i.e. when the number marker assumes the function of an article). The only structural pattern motivated by this factor is {N,NUM,NSG} which developed from the pattern of {N,NSG}.

10.3.4 Summary

In sum, we have observed that change in CNNCs involves counting and non-counting matters, and sometimes both together. In any case, it can be noticed that the change in CNNCs is motivated by most areas of grammar; these include phonetics (e.g. the fusion of numeral and numeral classifier), phonology (e.g. the loss of case in Irish), morphology (e.g. the loss of inflections in creoles and pidgins), syntax (e.g. loss of agreement in Arabic) and semantics (e.g. the special meaning of the dual marker in Ngiyambaa; or the emergence of the numeral classifiers in Chinese and Ejagham).

10.4 How did CNNCs evolve?

Viewing the evolutionary development of CNNCs globally, we can observe that both $CNNC_{SG}$ and $CNNC_{NSG}$ show two characteristics of an evolutionary nature. The first is that the evolution can be regarded as *bi-directional* or *reversible* in terms of structural complexity. At the initial stage, the constructions are simple (i.e. consisting of just two core constituents), and then the constructions show a progressive change towards complexity (i.e. consisting of more than two constituents). For example, $\{N, NUM\}$ turns to $\{N, NUM, CLF\}$; and $\{N, NSG\}$ turns to $\{N, NUM, SG\}$ or $\{N, NUM, NSG\}$. However, some constructions have then changed back to simple forms again. For example, $\{N, NUM, CLF\}$ may turn to $\{N, NUM\}$; and $\{N, NUM, NSG\}$ may turn to $\{N, NUM\}$. So, the patterns are subject to change to more complex forms and to more simple forms.

Another major characteristic of CNNCs is that some patterns are predictable and regular (i.e. several languages regardless of relatedness may take the same pathway) but others are *unpredictable* or *idiosyncratic*. Based on the analysis carried out in §10.1 and §10.2, some patterns of change may be predictable and can be established as general tendencies. The predictable and regular cases have to do with grammaticalization and lexicalization, for example, $N+$ ‘one-ness’ or ‘two-ness’ will develop into $\{N, NUM\}$ over time when there is a pressure to express the exact numbers. In addition, the pattern of $\{N, NUM\}$ where numerals are small numbers may turn to $\{N, NSG\}$ through the process of grammaticalization. In addition, $\{N, NUM, CLF\}$ can develop into $\{N, NUM\}$ when the numeral classifier is fused with the numeral. This phenomenon is not unusual in the numeral classifier languages. Besides, some cases do not involve grammaticalization and lexicalization. For example, the patterns of $\{N, NUM\}$ and $\{N, NUM, NSG\}$ may change to $\{N, NUM, CLF\}$ and $\{N, NUM, CLF, NSG\}$ respectively when the languages of $\{N, NUM\}$ and $\{N, NUM, NSG\}$ are in contact with numeral classifier languages.

However, many patterns of change take place in only one language or one group of languages for a particular reason. For example, the pattern of $\{N, NUM, OBL, SG\}$ took place in Russian because the dual marker was reinterpreted as genitive singular. As

far as the current data are concerned, such a reinterpretation has not been found to happen in other languages in which the dual marker and genitive case marker are found. The pattern of {N,NUM,OBL,SG} where the oblique is in the partitive case took place in Finnic languages when Finnic languages came into contact with the languages with {N,NUM,OBL,NSG}, where the oblique is in the genitive case. Such a developmental pattern is unusual. Besides, some of the changes in CNNCs are unpredicable because the changes in CNNCs depend on some other unpredictable changes, such as the decline of grammatical features, such as the loss of case/number in Goidelic languages, noun class in Kana, and agreement in Arabic dialects.

10.5 Conclusion

In conclusion, this chapter has illustrated the evolutionary trajectories of CNNCs along with the motivations for their historical development. With the comparative method and diachronic approaches, we can observe the evolutionary stages of CNNCs. It is conjectured that the construction consisting of a noun plus a word with a numerical interpretation (such as the words meaning 'alone' or 'company') may represent a possible initial stage of CNNCs. From that stage onwards, CNNCs have split into several types over time. The development is reversible in terms of structural complexity, and idiosyncratic in some cases. Besides, the contributory factors in the development of CNNCs involve a quantifying function, a non-quantifying function, and a mixture of both. These motivations for the evolution of CNNCs can be attributed to most areas of grammar. Also, some historical developments are regarded as contact-induced change.

11 The True Nature of Language: A Perspective from the Evolution of CNNCs

[T]he fundamental problem for linguistic theory is to understand (explain and predict) how linguistic structures evolve—come into being and change into new (sub) systems—and thereby to learn what the true nature of language is (Bailey 1982: 25).

So far, we have seen how CNNCs have come into being and have changed into new structural patterns. The final step, which is the ultimate goal of this current research, is to answer the question of what we have learned about the *true nature* of language from the study of the evolution of CNNCs.

Scientists understand the nature of life on Earth through evolutionary biology. Although language is not a biological organism, not even a concrete entity actually, it shares the *essence* or fundamental property with organisms. That is, *it can be transmitted with alterations*. Due to this fundamental property, *things evolve*. Therefore, both living things and language are instances of the same evolutionary paradigm which cover both the biological domain and the cultural domain (including language). The concept of a generalized evolutionary paradigm has been present in several places (e.g. Stevick 1963; Dawkins 1976; Lass 1990: 96; Croft 2000: 11-12). On analogy with genes, Dawkins (1976: 192) describes the cultural units which can be *replicated* or copied as “memes”; and linguistic units are accordingly referred to as “linguemes” by Croft (2000: 28). Evolutionary biology as a longer established science is therefore often employed as an analogy to understand and explain linguistics, especially in the field of language change since it provides a handy set of evolutionary explanations.

Frequently, an analogy is used in reasoning when we want to understand unexplained things. However, it has long been recognized that the application of

biological analogy to language is somewhat problematic. This is because biological organisms and language are not exactly structurally parallel. Despite this, it is generally accepted that the biological analogy to language change is indeed *possible* if we know how to use the biological concepts properly and cautiously (McMahon 1999: 314).

In this current chapter, this possibility will be examined by comparing the two evolutionary systems to see their similarities and differences. Linguistic data taken mainly from the investigation of CNNCs are used as evidence for our arguments. The comparison is expected to show that evolutionary biology provides a convenient *unifying* approach to help understand some parts of the nature of human language. The mismatches between the two evolutionary systems are believed to reveal the *distinctive* nature of human language.

This chapter is divided into three sections. First, the thesis provides a brief introduction to analogy (§11.1). Next, the misuses and sensible uses of evolutionary biology to language are reviewed (§11.2). Finally, a set of analogies between the two evolutionary systems are drawn and the mismatches are discussed (§11.3).

11.1 Analogy: What and Why?

Analogy is generally defined as a cognitive process of inferring an unexplained situation by relating it to a similar situation which is more familiar or better understood. An analogy is thus composed of two domains: one is the unexplained situation and the other is the familiar situation. The unexplained situation is known as *the target* because it is a situation or problem that we aim to understand, solve or explain to others. The more familiar or better understood situation is known as *the source* because it is a situation that is used as the basis for understanding or solving the problematic situation as well as explaining it to others. The two situations being compared must be similar in terms of a *relational structure*, i.e. the relationship that holds between the objects in each domain (Gentner 1983; Holyoak and Thagard 1997). Consider the simple arithmetic analogy taken from Gentner (1983: 156), $3:6 :: 2:4$. The similarity between the first domain (i.e. 3:6) and the second domain (i.e. 2:4) is the relationship that holds between the numbers in each domain; that is, “twice as great as”. To put it simply, when one is thinking analogically, it means that s/he is applying the relational structure in the source to the relational structure in the target.

According to Day and Gentner (2007: 39), the complexity of analogical thinking is due to two major cognitive processes: *categorization* and *schema activation*. One classic example is an analogy between the system of the atom and the solar system (Gentner 1983). With this example, we may analyse the processes of analogical thinking as follows.

Suppose that a person encountered an entity never experienced before i.e. the atom. Reacting to this unfamiliar entity (the target), (s)he categorizes the atom in relation to the previous similar entity that (s)he was familiar with (the source), that is, the solar system. This categorization leads him/her to activate the schema for the whole relational structure of the solar system and (s)he then applies the relational structure to the atom as schematized below:

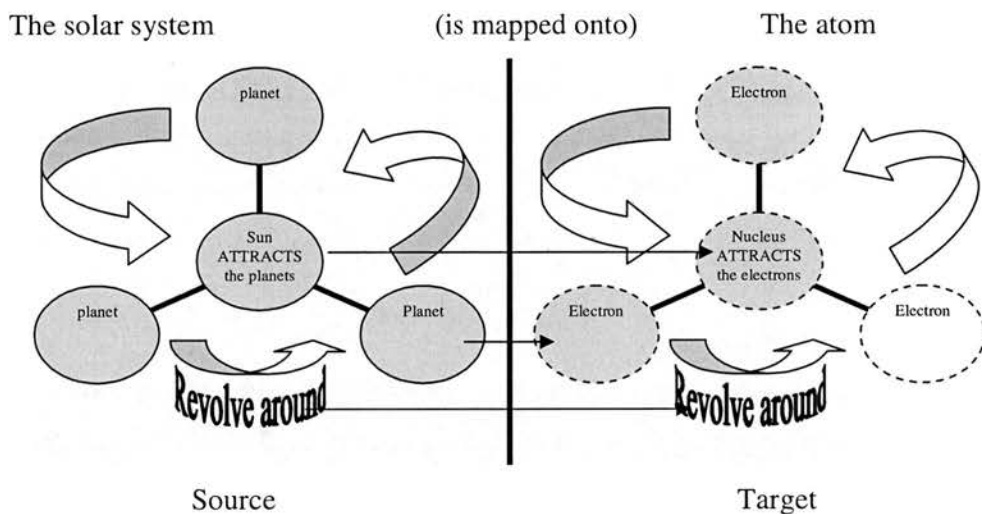


Fig. 11.1 Schema of analogy between the solar system and the atom

The atom (the target) is categorized in relation to the solar system (the source) because the atom is analogous to the solar system in terms of being a *central force system*. The cognitive process in the analogical mind at this stage is referred to as *categorization*. The process of categorization then activates the schema for the objects involved and the relationship between them; that is, the sun attracts the planets, causing the planets to revolve around the sun. The concept in the source domain is then transferred to the target domain. That is to say, the objects in the source and the target are arranged into a one-to-one correspondence. The transferral from the source to the target is represented by the thin arrows as shown in Figure 11.1. In the target domain, the objects involved are represented by dotted circles because they are not real yet, but the things to be realized. After mapping *the sun* onto *the nucleus*, and the *planets* onto the *electrons*, it is then inferred further that the nucleus attracts the electrons just as the sun attracts the planets; the electrons revolve around the nucleus just as the planets revolve around the sun (Gentner 1983: 159).

Immediate questions may arise: why is the sun mapped onto the nucleus, not the electrons; and why are the planets mapped onto the electrons, not the nucleus? This is

because the sun and the planets have an abstract (or generalized) property in common with the nucleus and electrons respectively. That is, the sun and nucleus are the central objects; and the planets and the electrons are the peripheral objects. The complete processes just described above illustrate what is generally called *analogical thinking*.

Since analogy involves transferring concepts across different domains, scientists have greatly benefited from it. Scientists use analogy to solve problems, explain new topics to others and make new scientific discoveries. For example, Ernest Rutherford discovered the central force system in the atom by using the analogy with the solar system (Dunbar 1999: 87); Johannes Kepler, by using an analogy with light, hypothesised that there was an invisible force from the sun causing the planets to revolve around it (Gentner 1999:17). So, analogy plays an important role in scientific reasoning.

However, one interesting point on the use of analogy is that an analogy need not require that the source and target domains be perfectly structurally parallel. Gentner and Kurtz (2006: 1) point out that:

Despite theoretical agreement on the importance of relational match, the empirical evidence is neither systematic nor definitive...Results show a consistent focus on relational matches as the main determinant of analogical acceptance. In addition, analogy does not require strict overall identity of relational concepts.

This point is noteworthy for the reason that the biological analogies to language indeed are rather imperfect. The comparisons are not as neat as the solar system analogy to the atom as just illustrated above.

11.2 Misuses and sensible uses of biological analogies

There have been numerous attempts to apply the concepts of evolutionary biology to language change since the nineteenth century. Some biological analogies and metaphors, especially those introduced into comparative and historical linguistics in the early period have come under heavy criticism. Recently, however, the parallels between biological and linguistic evolution which are mostly based on (post-) Darwinian theories seem to be more widely acceptable and regarded as fruitful. In this section, two frequently cited problematic biological analogies will be reviewed: *languages as organisms with life-cycles* and *languages as species* to illustrate the misuses of biological analogies (§10.3.1). Then, research works employing the (post-) Darwinian evolutionary concepts of *adaptation*, *natural selection* and *exaptation* will be reviewed to illustrate the recent acceptable uses of biological analogies (§10.3.2).

11.2.1 Misuses of biological analogies

11.2.1.1 Languages as organisms with life-cycles

A biological analogy which has been repeatedly attacked is the view that languages are looked upon as organisms having life-cycles—that is, they *are born, grow up, become old* and *die out* as evidenced by the quotation below.

Languages must be taken as organic natural bodies which form themselves according to definite laws, develop carrying in themselves an internal life principle, and gradually die, since they do not understand themselves any longer and shed or mutilate or misuse... (Bopp 1836: 1, translated by Morpurgo Davies 1987:84).

The fact that languages begin with their emergence, persist for a certain period of history and are no longer spoken may lead one to conceive of languages as organisms having life-cycles. For example, Sanskrit must have emerged at some point in India in the past. Then, it was used for some length of time and finally is no longer used nowadays. The claim that languages can be treated like *an individual's life* is understandable.

However, the analogy is true only in the sense of *language form* (e.g. English is a language form which is different from Old English) but not in the sense of communicative efficiency. Organisms undergo change over time not only in their forms, but also in their efficiency for survival. Therefore, for organisms, we can indicate the stages of life objectively through their physiology. For example, the reproductive ability suggests a stage of adulthood; the defects of organs such as worsening eye-sight are signs for a stage of ageing. The change in organisms at the stage of ageing can be absolutely regarded as *negative*. Whereas, for languages, there is general agreement that their communicative efficiency at any stage cannot be evaluated as better or worse than at the other stages. Because communicative efficiency never changes, it is impossible to indicate at what stage the language is.³⁶

According to McMahon (1999: 320-322), those who proposed that languages are like organisms undergoing change retrogressively such as August Schleicher (1821-1868) would claim that the old languages such as Sanskrit, Latin, and Ancient Greek represent the stage of fully developed languages and the modern languages descended from these three languages represent the stage of decay. Their proposal is based on the assumption that the morphological complexity of a language is an indicator of growth and decay. However, the loss of morphological complexity does not mean that the language loses communicative efficiency, and thus it cannot be used to indicate language decay. In some modern Indo-European languages where grammatical cases have been lost, a fixed word order may arise as a compensatory mechanism (Gil 2006: 93). There is no good reason to claim that the use of cases is better than the use of fixed word order, for instance.

One may also argue that, like organisms, languages finally will end with death as evidenced by the number of dead and dying languages all over the world. However, in this case, the death of languages involves the socio-political value, not because of their “inherent defects” (McMahon 1999: 322), such as lacking expressivity. A language may die because the speakers in that speech community use more prestigious languages. The

³⁶ The only sign that a language is near death is a small number of speakers. However, this is a matter of language form, not a matter of efficiency of communication.

Aboriginal people of Australia have adopted English in place of their native languages since English has become more beneficial to them for the new European way of life, such as shopping in supermarkets, studying in schools or other opportunities of advancement.

Apart from the view that languages evolve in a retrogressive direction, some scholars propose that languages are in fact moving in a progressive direction. Jespersen (1922: 324), for instance, argues that

That language ranks highest which goes farthest in the art of accomplishing much with little means...my formula contains two requirements: it demands a maximum of efficiency and a minimum of effort [...]

So, taking Jespersen's view, we can say that the modern Indo-European languages which have fewer grammatical categories (e.g. tenses and genders) than their parent languages should be regarded as better than their parent languages. These grammatical categories are semantically redundant and require the speaker to use more energy to learn as well as to communicate. So, the modern languages are better because they consume less energy in communication.

Although Jespersen's criterion to evaluate language is focused on communicative efficiency which can be compared with survival efficiency, it is not quite right to conclude that the modern languages are more efficient. Again, the fact is that languages at any stage are supposed to be equal in terms of efficiency. The speakers of some modern languages may not need to memorise complex case systems, instead they need to be aware of word order. The speakers of Latin might have needed to memorise cases but they might not have needed to be so aware of word order. Also, in fact, there are a great number of instances where languages evolve toward complexity and may seem difficult to learn. For example, the rise of numeral classifiers in Chinese results from a change in the conceptual meaning of nouns, rather than from communicative advancement.

In sum, in organisms, the survival efficiency will become worse when the organisms are becoming older, so the change can be evaluated as *retrogressive*. In

languages, the communicative efficiency remains unchanged over time, so the change cannot be evaluated in this way. The difference makes the biological analogy just mentioned above unsuccessful in this respect.

11.2.1.2 Languages as species

The idea of viewing languages as biological species is present in the language-family-tree model pioneered by August Schleicher in mid 19th century (McMahon 1999: 319). The family tree model is a diagram illustrating a group of languages sharing a common ancestor.

The fundamental idea is that languages gradually change over time, and once two speech communities have been separated for some length of time, these languages will split up into new languages. The divergence of language corresponds to the speciation of biological species (i.e. the splitting of a species into new species).

The family tree model has been widely used in historical linguistics, right up to the present day. Nevertheless, this model has been considered problematic (e.g. Lehmann 1962: 139-142; Dixon 1997: 29). The major problem is that biological species cannot normally interbreed with each other (Charlesworth and Charlesworth 2003:9) but languages can be mixed due to contact with other languages no matter how genetically distant they are. So, some modern languages are in fact products of language contacts; not direct descendants from parent languages. For example, English is affiliated to the Germanic branch and is accordingly genetically related to modern German. English is thus supposed to be similar to German. In reality, however, English has been influenced by some other languages in its history, especially French (as evidenced by the sizeable French vocabulary found in English, as well as by some aspects of grammar and phonology), which is affiliated to the Italic branch. So, English is in fact a product of Germanic and Italic languages at least (Dixon 1997: 52) as illustrated in the diagram below.

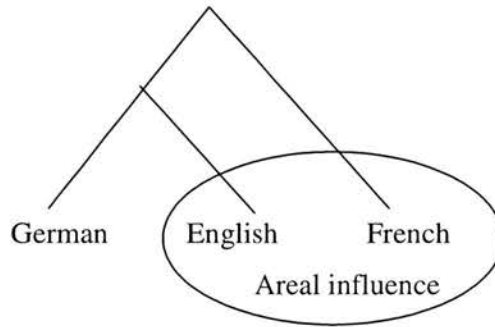


Fig. 11.2 Phylogenetic schema: German, English and French (Dixon 1997: 52)

However, the language-family-tree model cannot show this fact because the model is based on the assumption of genetic relationship only; and it does not deal with language contact at all. For this reason, the family tree model therefore cannot be applied to the mixed languages like pidgins and creoles in which the modern language systems are the result of the mixture of superstratum and substratum languages (cf. Thomason and Kaufman 1988: 1-3). As previously mentioned, interbreeding is not normally found between different biological species, but a process perhaps equivalent to interbreeding may be found in language contact.³⁷

To conclude, according to the differences in certain features between biological organisms/species and language just described, the application of a biological analogy to linguistics has been strongly rejected by some in the linguistic community, among them, Ferdinand De Saussure (McMahon 1999: 314) and Edward Sapir (Stevick 1963: 159). Rather, they argue that language has its own nature and should be studied in its own right, as seen from the quotation below.

We now view language as a set of social conventions so complex that a simple biological or geometrical model is totally inadequate. Rather than force one on language, we attempt to understand it in its complexity (Lehmann 1962: 142).

³⁷ Due to the problem of the family tree model, Johannes Schmidt (1872 in Lehmann 1962: 140-141) proposed the *wave theory* where contact-induced change can be shown in the model. However, the model fails to depict the historical development of a group of languages (see Lehmann 1962: 140-142, for more detail).

11.2.2 *Sensible uses of biological analogies*

In this section, a few previous works applying evolutionary biological analogies to language are mentioned. These analogies are all based on (post-) Darwinian evolutionary ideas.

Darwinian evolutionary theory refers to the biological evolutionary theory introduced by Charles Robert Darwin (1809-1882) who is the founder of the theory of evolution by natural selection. The central idea of the theory is that all living things on Earth share a common ancestor. The variation of living forms is a result of their adaptation to their environment. Those who fit or are able to adapt to the environment will survive and reproduce offspring. Evolutionary ideas developed from Darwin's theory at a later time are known as *post-Darwinian* or *Neo-Darwinian* theories.

Several Darwinian evolutionary ideas have been transferred into linguistics to explain diachronic change in language. Two closely related key ideas in the Darwinian evolutionary framework (which are mentioned very often in evolutionary biology and linguistics) are *adaptation* and *natural selection*. The term "adaptation" is generally defined as a useful characteristic that organisms have. These useful characteristics help the organisms to survive and reproduce in their environment. Those organisms which cannot adapt to their environment will die out while those which can will survive and have reproductive success. This means that they are (naturally) selected and so the process is referred to as *natural selection*. The two concepts have been metaphorically employed to explain the underlying mechanisms of language change. To illustrate, a well-known issue such as the universals of colour terms (Berlin and Kay 1969) will be re-explained in terms of adaptation and natural selection.

11.2.2.1 Darwinian ideas: adaptation and natural selection

Berlin and Kay (1969) propose the implicational hierarchy of basic colour terms³⁸ as shown below.

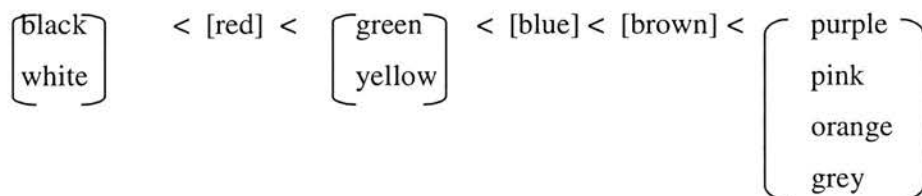


Fig. 11.3 Implicational hierarchy of basic colour terms (Berlin and Kay 1969: 4)

The hierarchical diagram shown above is supposed to be exemplified in all or almost all of the world's languages. It means that if a language has the colour terms on any position to the right, it will also have all the colour terms to the left. For example, if a language has the colour terms for green or yellow, it will also have the colour terms for black, white and red; the language may or may not have other colour terms to the right (e.g. blue or brown). In this case, it can be observed that the colour terms for black and white are ranked the highest in terms of the possibility of being found, whereas the other colour terms are ranked lower and lower respectively through the hierarchy from left to right.

Dik (1997: 30-33) observes that Berlin and Kay's universal implications also suggest that the *frequency of occurrence* of colour terms will decrease through the hierarchy from left to right. The idea is applied either to the same language or cross-linguistically. For example, the colour term for red is supposed to be used more frequently compared to other colour terms to the right (e.g. green, yellow) within the

³⁸ Berlin and Kay (1969: 5-6) have given several criteria for defining *basic colour terms*. For example, the basic colour terms refer to the colour terms that are generally known in the speech community. Also, they are not built up from the lexemes already in the language, so their meanings cannot be predicted from their components (e.g. *lemon-coloured* is not a basic colour term because we can predict the colour from the word *lemon*). Besides, the basic colour terms are usually used for naming the colours that are regarded as psychologically salient in the mind of the users of the language (i.e. the colour terms that the speaker would think of firstly when asked to list some).

same language. When viewed across languages, the colour term for red is supposed to be found in more languages than that for green. Dik (1997: 33) further suggests that the frequency of occurrence of the basic colour terms either in one and the same language or across languages can be explained with reference to *functional pressures* in language use. This is because some colours are needed for use very often, and they therefore have more chance to be assigned a particular name—they are functional in this sense.

Haspelmath (1999: 6, 15-16) points out that the basic colour term hierarchy can be viewed with reference to *adaptation*. The central idea of the notion adaptation given by Haspelmath is that “Grammatical structures are adapted to the needs of language users”.³⁹ The adaptive explanation in Haspelmath’s sense includes the notions of *adaptation* as well as *natural selection*. That is to say, the basic colour terms that already exist in the language are less likely to be lost from the language. This is because they have been used to refer to the colours so frequently that the terms have been established in the cognition of the users. The fact that the basic colour terms are frequently used suggests that they must be useful or must respond to the needs⁴⁰ of the users. In other words, they have a certain adaptive characteristic for the users. Such an adaptive characteristic can be compared to an *adaptation* in the biological domain, assuming that the basic colour terms are compared to living organisms, and the users are mapped to the environment which the living organisms inhabit. Like organisms which can adapt well to the environment, those basic colour terms will be selected to be established firmly in the user’s lexicon and are less likely to be lost from the language. This clearly corresponds to the survival of organisms; and the process is comparable to *natural selection* in the biological domain. Therefore, it can be claimed that the universal hierarchy of colour terms as present in the world’s languages evolves to fit communicative needs.

There is another interesting point mentioned in Dik (1997: 32) which should be highlighted with respect to adaptation as well. Dik predicts that if the colour term system

³⁹ Of course, *adaptive explanation* is not the only approach to explaining the diachronic development of grammatical structures. See Comrie and Kuteva (2005: 185-207) for a non-adaptive explanation as an alternative approach to the historical origins of synchronic grammatical structures.

⁴⁰ The colour terms are needed in the sense that they help to refer to the colours that are salient psychologically in the users’ minds. The colours vary in degrees of cognitive salience. For example, the black/white colours are more salient than the red colour.

of a language decays, the colour terms to the right of the hierarchy tend to get lost earlier than those to the left. For example, the colour term for blue is likely to get lost before that for red. Such a constraint on diachronic change can be explained in terms of adaptation and natural selection as described in the preceding paragraph. In the biological world, the organisms that can be adapted most will be selected and consequently, they will be the last ones to die out or become extinct. In linguistics, we can predict that the grammatical feature which is the most important or useful to users will be ranked highest in the hierarchy and will then be selected (as evidenced by most frequent use). The most important (frequent) grammatical feature will be the last one to undergo change or disappear. Retrospectively, in the biological world, the organisms which do not adapt well to their environment will die out or become extinct first. In linguistics, it can also be predicted accordingly that the grammatical features which do not adapt well or are the least beneficial will be lost from the language system first. For example, based on the hierarchy of colour terms, the colour terms for pink, purple and grey, which are ranked the lowest, would be supposed to be lost from the system earlier than the colour term for blue.

The principle just described above (i.e. on the correlation between frequency of occurrence and the rise/decay of grammatical features) may also be applied to other grammatical hierarchies, such as the numbers (e.g. singular > plural > dual > trial; Greenberg, Corbett 2000: 38⁴¹), the vowel systems (/i, a, u/ > /ɨ, ε/>, /ɔ/ etc; Crothers 1978:136⁴²); or lexicon (e.g. the basic terms⁴³). With the principle based on the evolutionary ideas of *adaptation* and *natural selection*, we can make a reasonable hypothesis about the chronological order of the rise and decay of grammatical features in language, using the hierarchy of occurrence as an evolutionary ladder.

⁴¹ See Corbett (2000: 38-50) for more detail.

⁴² See de Boer (2001) for an evolutionary perspective on the origins of vowel systems.

⁴³ The basic terms refer to the terms which are significant and frequently used in the everyday life of the speakers in the speech community, for example, the terms denoting numbers (e.g. one, two), pronouns, and kinship terms (see Holmes 2007: 1-2 for research works supporting the adaptive explanation).

11.2.2.2 Post-Darwinian idea: exaptation

Not only original Darwinian concepts have been borrowed into linguistics. Neo-Darwinian or Post-Darwinian concepts have also been taken. One of these is the concept of *exaptation*. We have learned earlier that the term adaptation refers to the organism's adaptive characteristics that help it to survive. Currently, biologists have pointed out that the term adaptation should be viewed more narrowly as "any feature that promotes fitness and was *built by natural selection* for its current role" only (Gould and Vrba 1982: 6, emphasis mine) So, if the historical origin of the features are not for its current role or do not evolve by natural selection, they will not be referred to as adaptation. Gould and Vrba (1982:6-7) refer to the features which evolve as a side-effect like this as *exaptation*. For example, bird feathers originally evolved for temperature regulation but were later applied for flight.

It is noted in Gould and Vrba (1982: 6) that the term *exaptation* is needed in evolutionary biology. That is, it would help the evolutionists not to assume mistakenly that every current feature evolves for its current role, as evidenced in a note in Bock's definition (1967: 63 cited in Gould and Vrba 1982: 6) below.

On theoretical backgrounds, *all existing features* of animals are adaptive. If they were not adaptive, then they would be eliminated by selection and would disappear. (italics mine)

Like evolutionary biology, the concept of exaptation is useful for historical linguistics in terms of the origins of grammatical features. A number of grammatical features originally evolve for a certain function, but due to some historical reason, they are later used for some other functions. The grammatical change in this way corresponds to the concept of exaptation in evolutionary biology. Lass (1990) gives several examples of exaptations used in historical linguistics. For instance, in Modern English, the subject normally precedes the verb, and the pronoun is obligatory, so we can know which argument is the subject of the verb. In the past, English had a much freer-word order than in the present and subject pronouns were not obligatory. In a situation like this, ambiguity was avoided through cases and the argument inflections. For example, the *-s* ending was suffixed to the verb to indicate that the subject of the verb was the third

person singular, and that the action was in the present tense form (Lass 1990: 99). This is actually another example of *adaptation* in language. However, the –s ending in some dialects (e.g. Reading English) is used in all person number forms as an indexical marker of vernacularity. It has therefore changed (or *exapted*) for sociolinguistic purposes. For example, *I starts Monday, so shut your face.; They calls me all the names under the sun, don't they?* (Lass 1990: 99; Cheshire 1982: 31).

To conclude, it can be observed in Section 10.3 that the use of biological analogy in the past was not quite successful because of the different nature of organisms and language in terms of efficiency and interbreeding. However, recently, linguists have turned to biological analogy again by looking at evolutionary mechanisms in (post-) Darwinian theories. The careful application of biological concepts in this way has become fairly widely accepted and regarded as useful (McMahon 1999: 315). In the next section, we will examine to what extent biological analogies can be applied in linguistics.

11.3 Analogies between organisms and linguistic units

In historical linguistics, there are a number of factors involved in language change. In most cases, changes in language are difficult to predict as evidenced by the study of CNNCs. Still, there should be some causal mechanisms and predictable outcomes that can be observed and generalized in language change. We may refer to the generalized patterns of change as the *nature of change*. The so-called *nature* in language change can be understood by analogy with (post-) Darwinian evolutionary concepts. The analogy, nonetheless, may be most practical when the application is not too restricted in terms of the mapped features between the two domains. Since linguistic units are psychological and socio-cultural artefacts which must be different from biological organisms, perfect mapping is obviously impossible. Instead of forcing the linguistic domain onto the biological domain and ignoring the mismatches, these mismatches should also be taken into account. This is because these mismatches in fact make human language distinctive in evolutionary systems, revealing the *true nature* of language. This will be illustrated in the analogies, as follows.

11.3.1 Evolution: descent with modification

Biological organisms are replicable units. That is, they can be copied and propagated. A historical process involving descent through genetic inheritance is a general definition of the term (*biological*) *evolution*. The descent, nevertheless, normally involves certain modifications and this gives rise to diversity, even though all organisms on Earth are presumed to descend from a common ancestor (Caldwell et al. 2006).

Linguistic units⁴⁴ are analogous to biological organisms in that they are replicable units. When they are learned and used by others or in later generations, we can say that they are replicated (Croft 2000: 3). Like organisms, when the linguistic units are replicated, they may be replicated with some change. Again, like organisms,

⁴⁴ The term “linguistic units” here refers to linguistic elements at any level. For example, /s/ is a linguistic unit at the phonemic level; CNNCs are linguistic units at the syntactic level.

the change over generations can lead to linguistic diversity and this means that language evolves.

In sum, the essence of the analogy between biological organisms and linguistic units is “the altered replication” (Croft 2000: 3) which corresponds to Charles Darwin’s “descent with modification” (Darwin 1859). Since linguistic units share this fundamental property with biological organisms, it is worth further exploring the mechanisms in biological evolution and seeing to what extent these mechanisms can be mapped onto those of language.

11.3.2 Origins

Biological organisms and language are quite similar though with a slight difference in this respect. Both biological organisms and linguistic constructions like CNNCs did not emerge simultaneously without foundation, rather they developed from pre-existing forms or an embryonic stage.

In biology, life on Earth developed from a stage where certain inorganic elements were composed; that is, a composition of chemical elements (not biological organic elements). We may refer to this stage as an embryonic stage. CNNCs are similar to biological organisms in this respect. Before the rise of CNNCs (i.e. before humans used numerals proper), there existed some lexical words which had the potential to develop into numerals. Examples are words meaning ‘alone’, ‘small’, and ‘face each other’. Since there was no category of numerals as such in language of that time, this stage should not be regarded as a cardinal numeral-noun constructional stage.

In biology, after certain chemical reactions, the chemical composition then changed into organisms consisting of only one cell (or *single-celled organisms*). The single-cell organisms represent the earliest and simplest forms of life. Some of these simplest organic forms turned into more complex organisms consisting of more than one cell (or *multicellular organisms*). Slightly different from biological evolution, the embryonic CNNCs have been shown *either* to evolve into the simple form (i.e. a noun plus a numeral) which are CNNCs proper, or the simplex form (i.e. a noun plus a non-singular marker; e.g. *boy-s*). The simplex form at this stage is not a cardinal numeral-

noun construction as such because it does not contain a numeral. This means that when the CNNCs were actually established, it would have been possible for the first form of CNNCs proper to be a complex form, as opposed to the simple beginnings which tend to be characteristic of biology. This can be illustrated by the case of {N,NUM,NSG} where NSG was grammaticalized from the words conveying plurality (i.e. Noun plus *many* > Noun plus NSG > {N,NUM,NSG}); illustrated by this artificial version of English as *dog-many* > *dog-s* > *two-dog-s*).

In sum, the origins of CNNCs are more varied than those of biological organisms. The thesis does not argue that the origins of CNNCs illustrated above definitely represent a scenario for language origins. However, it suggests that the origins of linguistic systems are a relatively complicated matter compared to the origin of living things.

11.3.3 Evolutionary trend

Evolutionary trend refers to a directional change that continues for some length of time so that the change can be predicted. If the evolutionary change is random or fluctuating, that course of change will not be referred to as *evolutionary trend*, rather it will be referred to as *drift* (Caldwell et al. 2006).

For biological organisms, at the earliest stages, the evolutionary trend exhibits direction toward complexity. All organisms presumably began with only one cell. Later, some of these organisms changed into more complex forms, consisting of more than one cell. After the earliest stages of life, the overall evolutionary trend is hard to see. This is because 90% of the biological organisms today are single-cell organisms, typically bacteria. Besides, some parts of the organs of living things have reduced or gotten lost, but other parts may become increasingly complex. For example, humans have lost their tail, but their brain has become increasingly more complex; and the hind wings of flies have evolved into halteres (i.e. complex organs for flight control). Moreover, there are several other pieces of evidence showing the trend toward simplicity. Insects evolved from arthropods with more than six legs. Birds and snakes have lost parts. Some stick insect lineages no longer have their wings. Wild oat flowers have been simplified. So,

the issue of the overall trend of evolution is not straightforward (Caldwell et al. 2006). Maynard Smith and Szathmary (1999: 15) have given a well thought-out conclusion on the evolutionary trend in organisms as:

The theory of evolution by natural selection does not predict that organisms will get more complex. It only predicts that they will get better at producing and reproducing in the current environment, or at least that they will not get worse.

As for CNNCs, after the constructions have become established in human language, they may become complex in terms of the number of extra elements. For example, {N,NUM} may turn into {N,NUM,CLF} or {N,NUM,OBL,SG}. The complex constructions may in turn become more complex. For example, {N,NUM,NSG} may become {N,NUM,NSG,CLF}. However, some complex constructions may become simpler. For example, {N,NUM,CLF} may become {N,NUM}. We can see that the evolutionary direction of CNNCs (and perhaps of some other linguistic structures) is somewhat random. So, in this case, the evolutionary direction of language may not be so different from the evolutionary direction of biological organisms.

In sum, the quotation given by Maynard Smith and Szathmary (1999: 15) just mentioned above is also true of language. If indeed we can define the complexity and simplicity in language, it is still impossible to predict that languages will evolve towards complexity or simplicity. This is because there would be a number of cases showing the reverse directions. Even in the same language, in different areas of grammar, the directional change can be different (McMahon 1999: 323). Thus, all that can be said is that human language will evolve to maintain the efficiency of communication. The changes can occur at all times due to several kinds of factors, a change in a certain area of grammar can affect the other areas, leading to lesser effectiveness or efficiency of communication. That is why language evolves to adapt itself infinitely.

11.3.4 Gradual change and sudden change

Language is similar to biological organisms in that it may illustrate *gradual change* as well as *sudden change*. For biological organisms, gradual change may show *transitional forms* (i.e. intermediate states between the ancestors and their descendants) or co-existence of two split forms over a period of time. For example, it is known from the fossil records that whales are a species which have gradually developed from terrestrial mammal ancestors. There are several modifications during the transitional periods. An example of a modification is found in the breathing apparatus. See the nostrils in Fig. 2 below. From the whale's ancestor (the left picture), the nostrils are at the front of the skull like most land mammals. Later, the nostrils are at the middle of the skull as shown in the centre picture. For the whales today (the right picture), the nostrils are at the top of the skull. The centre picture illustrates the transitional form. This change resulted from the adaptation made from the terrestrial environment to the marine environment.

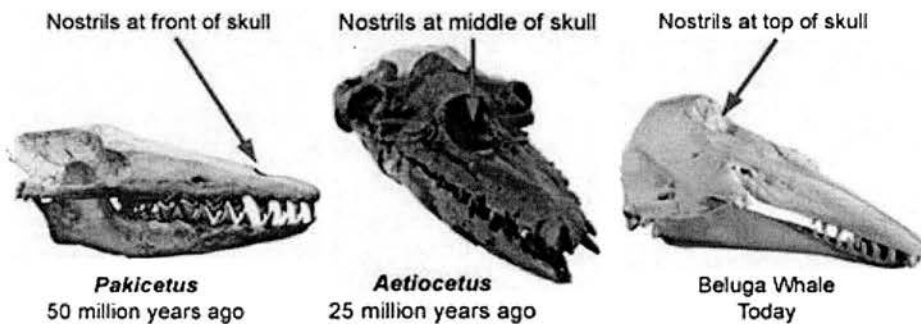


Fig. 11.4 The skulls of the Beluga Whale and its ancestors in different periods
From <http://evolution.berkeley.edu/evosite/lines/1Atransitional.shtml> © 2008 The University of California Museum of Palaeontology, Berkeley, and the Regents of the University of California (Used with permission)

Biological organisms can also show a sudden change. This idea is known as *punctuated equilibrium*. This theory is first proposed by Eldredge and Gould (1972 in Levinton 2001: 311). The central idea is that a change in species happens rapidly, say within thousands of years, as a result of catastrophic changes occurring periodically in the environment, such as floods or droughts (*punctuation*). Species need to adapt abruptly to a new environment, giving rise to speciation (i.e. a split of new species).

After the bursts of new species, there will be a very long period of stability or stasis (*equilibrium*). During the stasis, species remain unchanged or minimally changed until the occurrence of the rapid evolutionary change again. This theory does not sound implausible but the main problem is that the theory is hard to prove with concrete evidence. The lack of fossil evidence for transitional periods does not necessarily mean the transitional periods do not exist.

Like organisms, language often changes gradually. There is a good deal of evidence of transitional forms as well as the co-existence of types. However, language sometimes suddenly changes in a relatively short period of time due to some punctuated event.⁴⁵ Both the gradual changes and sudden changes can be found in CNNCs.

In Chinese, the evidence from historical documents suggests that numeral classifiers gradually developed from {N,NUM} and it took more than 3,000 years to complete the process (Wang 1996). The gradual change is evidenced by the transitional forms and the co-existence of forms. It seems likely that the transitional form between {N,NUM} and {N,NUM,CLF} is the use of repeaters (e.g. 'ox'₂ in *ox₁-one-ox₂*), during the time when there was no numeral classifier proper. Also, during the gradual change, there are two constructions ({N,NUM} and {N,NUM,CLF}) present at the same time although {N,NUM,CLF} at the beginning is not widespread.

However, during the time of gradual change, there was also a sudden change. In the Han period, the numeral classifiers had increased remarkably from 15 in the late Archaic period to approximately 50 in this period, and the increase of numeral classifiers might have accelerated the spread of the {N,NUM,CLF} construction. Another example of sudden change in language can be found in Pidgins and Creoles where CNNCs which spread from the superstrates have been changed into {N,NUM}. In this case, the punctuated event involves economic and political matters.

In sum, biological and linguistic forms have been known to change gradually. In the biological world, sudden changes might have happened periodically at some point in

⁴⁵ See Dixon (1997) for the application of the theory of punctuated equilibrium to the origin and evolution of language.

the long history of life. Unfortunately, due to the lack of evidence, this theory remains controversial. However, in language, sudden change may happen due to changes in politics or economics.

11.3.5 Speciation

Speciation is a term used in biology referring to a splitting-event of species in a lineage, giving rise to new species. This can be illustrated by branching points in the phylogeny (i.e. a tree diagram of life) below. In the phylogeny of the fruit fly below, there are two speciation events at the branching point marked by the circles. For example, at some point of time, due to genetic changes, the common ancestor of the fruit fly *Drosophila dentissima* and *Drosophila melanogaster* split up into two separate lineages.

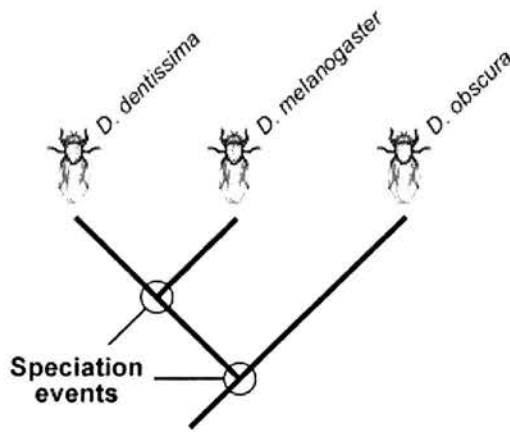


Fig. 11.5 Speciation events in the partial *Drosophila* phylogeny
From <http://evolution.berkeley.edu/evo101/VBDefiningSpeciation.shtml> © 2008 The University of California Museum of Palaeontology, Berkeley, and the Regents of the University of California (Used with permission)

How and why did the speciation take place? Scientists point out that a physical barrier (e.g. mountains, rivers) is a common way of preventing different groups of the same species from mating with one another. At this stage, speciation has not yet taken place. Each population evolves differently over the course of many generations due to different *selective pressures* such as ecological conditions. Geographic isolation may

lead to the process of *reproductive isolation* at a later stage. That is, the two populations evolve so differently that their genetics are different and they cannot interbreed to produce viable offspring. At this stage, speciation has occurred.

Let us turn now to language. CNNCs also split up into many types mainly because of reproductive isolation akin to biological speciation. For example, in Indo-European languages, the CNNCs split up into several types as shown in the figure below.

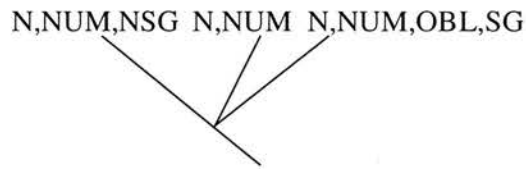


Fig. 11.6 Splitting event of a partial phylogeny of Indo-European CNNC_{NSG}⁴⁶

One reason is that several groups of the ancestors of the Indo-Europeans had emigrated far from their homeland (presumably somewhere around central Asia), and as they were separated they experienced different selective pressures. For example, in Goidelic languages such as Irish and Scottish Gaelic which belong to the far-flung Western branch of the IE family, the structural pattern of {N,NUM} (rather than {N,NUM,NSG}) is widely used for CNNC_{NSG}. This is because the languages experienced the phonetic erosion of final unaccented syllables, giving rise to the loss of number marking. At the Eastern branch of IE, in some Slavic languages, the dual was reinterpreted as the genitive singular marker because the two forms were identical in some nouns. The changes as happened in Goidelic are different from those in Slavic languages because each has experienced different selective pressures or drift. The two cases show reproductive isolation (in the sense that each has developed their own CNNCs differently).

⁴⁶ The diagram as shown above is drawn by analogy with the *phylogenetic starburst* in biology. The phylogenetic starburst is used when the chronological order of speciation cannot be indicated due to rapid speciation. Here, it means that {N,NUM} does not develop from {N,NUM,OBL,SG} and vice versa due to rapid divergence.

Reproductive isolation in biology results in speciation (i.e. the two species which used to share a common ancestor can no longer interbreed even if the physical barrier breaks down). However, this is not the case for language. It seems likely that the Celtic CNNCs and the Slavic CNNCs would influence each other to some extent if they happened to be in contact over generations, like most cases where the speakers of different languages (or even different families) have interacted closely over a long period of time. For example, Finnish which presumably had {N,NUM} changed to have {N,NUM,OBL,SG} when the language came into contact with the Baltic languages. Nonetheless, there perhaps exist some constraints on the mutual influence between languages with regard to CNNCs.

11.3.6 Mechanisms of evolution

The mechanisms of biological evolution explain the vast diversity in organisms. These explanations can be borrowed to account for the change in language. The mechanisms of evolution include *mutation*, *gene flow*, *genetic drift* and *natural selection*.

11.3.6.1 Mutation

Mutation is a change in an organism's genes, accidentally caused by some errors occurring during replication. Mutation produces new variant forms in the population. For example, a mutant form of the fruit fly *Drosophila melanogaster* had a leg at its head in place of its antenna (Charlesworth and Charlesworth 2003: 6). Most mutations have *negative effects* to the mutant. This is because the organ which has well adapted to the environment over a long period of time is changed when the mutation occurs, and so this is likely to be dangerous to the mutant. On the other hand, if the environment changes, mutations may prove advantageous, increasing chances of survival and reproduction, and may finally replace the non-mutant forms.

In linguistics, the term “mutation” is used in a broad sense to refer to the process in which a replicated form of a linguistic unit is altered by speakers.⁴⁷ In this sense, linguistic mutation corresponds to the traditional term *variation* as used in the news headline “Language ‘mutations’ affect least-used words” (NewScientist.com 11 October 2007, Bob Holmes)

Holmes (2007) uses the term “mutation” to refer to an innovative variant form in discussing the change in less frequently used words. In the news, Holmes reports two studies, one by Mark Pagel and colleagues and the other by Erez Liberman and colleagues. According to his report, the two studies reveal that “the most frequently used words tend to remain unaltered, while rarer words are more likely to change.” For example, the word representing the concept of the number ‘two’ is unlikely to change (here change means using a new word, not just a phonological change) as evidenced by the fact that the Indo-European languages use the same cognate forms for number ‘two’. Another example lies in the history of English, where the verbs that are used very often in everyday life resist regularization (here it means using the *-ed* suffix instead of irregular forms). On the contrary, the verbs that are not used very often tend to be regularised. So, in this sense, mutation may correspond to variation.

However, due to the fact that biological mutation involves erroneous replication happening accidentally, it is argued that the variant forms which are functionally motivated (the use of a linguistic form in a special context or for ease of learning) should not be referred to as mutation. The regularization of the irregular verbs in English may illustrate ease of learning; that is, the regularization helps the learners to acquire the system with less difficulty. In this case, the alternation should be referred to as an adaptation rather than a mutation because the system adapts itself to favour its users. The term linguistic mutation will be reserved for an alternation caused by mistaken learning or an alternation with no underlying functional explanation for its occurrence, such as

⁴⁷ In a narrow sense, the term ‘mutation’ is used in phonology, referring to the process in which “In common with the other Celtic languages, some of the initial consonants of Welsh words vary according to their grammatical context, for example: *cath* ‘cat’; *ei gath* ‘his cat’; *fy nghath* (‘my cat’)” (Ball and Müller 1992: 6). This is a different sense of ‘mutation’.

the minor change in a Standard Thai tone from falling tone to falling-high tone in the young generation. This change cannot apparently be attributed to any functional explanation.

Turning to CNNCs, it is hard to find evidence for mutation. This is perhaps because most changes can be attributed to functional explanations or language contact (which of course cannot be described as *mutation*). However, an instance is found in Russian {N,NUM,OBL,SG} where some nouns in the dual forms were similar to the *genitive singular* form and then the dual was reanalysed as the genitive singular forms. This case is considered as an instance of erroneous replication; that is, the dual is mistakenly changed to the genitive singular, making {N,NUM,DU} become {N,NUM,OBL,SG} without any underlying functional explanation.

Regarding mutation, language and organisms may be similar in the way that both possess mechanisms causing variant forms in the population. However, organisms and language may be different in certain respects. First, for organisms, a mutation with advantageous effect to the mutant is extremely rare, as most are negative (Houle and Kondrashov 2003: 5-6), while in language, the variant forms tend to survive, and perhaps end up replacing the non-mutants. Second, the mutation rate in organisms is very low. For example, the mutation rate in mice is 46: (over) 1,000,000 (Russell and Russell 1996 in Houle and Kondrashov 2003: 2) whereas in language, mutation is quite common, especially in phonetics and semantics (e.g. using a word with incorrect meaning). Mutation in organisms will be passed on from parents to offspring only if the mutation occurs in reproductive cells (except asexually reproductive organisms like bacteria where this principle is not applied). On the other hand, in language, a mutation is easily passed onto the next generation and mostly affects the language system.

11.3.6.2 Gene flow

In biology, gene flow will occur when genes from one population are transferred to another (normally in the same species). For example, westerners may move to the eastern world and marry the local people and have children. The genetic material of the

westerners will be transferred to those children, giving rise to genetic variation in the receiving population.

In language, *gene flow* is comparable to the linguistic phenomenon known as *language contact*. There are several interesting similarities and differences between gene flow and language contact (cf. §3.2.2.2).

It can be observed that gene flow is similar to typical language contact; that is, when different populations (speech communities or biological populations) are in contact, a replicable unit (linguistic material or a gene) from one population is transferred to and spread in the new population, leading to variation and some change in the population. In biology, gene flow helps to reduce the genetic variations in species because of the combination of two gene pools. This will reduce the chance of speciation (i.e. the split of species). Similarly, in language, language contact acts against the birth of a new language developed from a dialect by recombining the two features of the two dialects. This is the opposite of the situation where a new language arises through geographical isolation.

Although language contact and gene flow may be similar in general, they are different in a few ways. First, in terms of constraints on the transmission of the replicable units, gene flow can occur freely between the populations of the same species *only* (or perhaps related species, in which case, the process is known as *hybridization*); not of different species due to genetic barriers. On the contrary, language contact may occur with far fewer restrictions. In language, linguistic borrowing is quite widespread, typically in terms of lexical morphemes, regardless of genetic, typological, or areal relationship. Grammatical borrowing is nevertheless not as common as lexical borrowing. Also, grammatical borrowing is likely between the dialects of a single language or languages with similar grammatical systems (Meillet 1921: 87 in Thomason and Kaufman 1988:14-15; Thomason 2006: 341). Givón (1979: 26 in Thomason and Kaufman 1988: 15) as a functional linguist, regarded the structural system as a non-arbitrary system (i.e. functionally/ semantically motivated) and so believed that borrowing foreign structural material would be unlikely. Bickerton (1981: 50) points out that because language has systematic structure, borrowing incompatible structural

features cannot happen. Although Givon's and Bickerton's argument might be slightly too strong and there must be many counterexamples (see Thomason and Kaufman 1988: 13-34 for examples of structural borrowing), their argument is probably true in terms of *natural tendency*. It is likely that the grammatical systems in a language have developed or adapted systematically over a considerable period of time. The foreign structural features which are incompatible with the well-developed grammatical systems are unlikely to be borrowed. Hence, in terms of constraints on grammatical borrowing, we can see that language contact is not so different from gene flow. That is, it is more likely when the two languages or populations are genetically related.

In any event, the difference between biological organisms and language perhaps lies in the degree of *flexibility* of receiving foreign features. Biological organisms are more resistant, whereas language as a socio-cultural artefact is more flexible in receiving foreign features. This characteristic of language is highlighted in Thomason and Kaufman (1988: 15). As mentioned earlier in §3.2.2.4, they strongly claim that the linguistic structural constraints can, in fact, be overcome by social factors such as "relative population sizes, length of contact and degree of bilingualism" (Thomason and Kaufman 1988: 65-66).

In terms of CNNCs, the pure structural loan such as the borrowing of {N,NUM,NSG} into the languages with {N,NUM} is not found. It seems to me that number borrowing is relatively rare compared to the spread of numeral classifiers. This is because numeral classifiers spread through the numeral borrowing which is more lexical-like, for example, the spread of {N,NUM,CLF} from Indo-Aryan languages to some Dravidian languages (which presumably had {N,NUM} and {N,NUM,NSG}). The {N,NUM,CLF} construction is commonly found in the northern group and central group of the Dravidian languages but is extremely rare in the southern group which might not have been under heavy influence of the Indo-Aryan languages. So, in this case, the grammatical structure in the Dravidian languages has changed because of lexical borrowing, not structural borrowing.

The second difference between gene flow and language contact is that language contact may give rise to the loss of linguistic material in the receiving language whereas

there is nothing lost in gene flow. For example, Arabana-Wangkanguru has lost the trial number due to the influence of English which is not marked for trial (Hercus 1994: 64). In addition, in gene flow, the genetic material can be clearly observed, but in language contact, the borrowed feature is not necessarily clearly seen. That is to say, a grammatical structure of the receiving language may produce a *new* grammatical structure which is not found in the source language (and was not previously in the receiving language). The new structure comes about by *combining* the grammatical structure of the source language and the receiving language. For example, in Finnish (which presumably had {N,NUM}), the {N,NUM,OBL,SG} construction might have arisen due to contact with Baltic languages where {N,NUM,OBL,NSG} is widely used. The current structure of {N,NUM,OBL,NSG} is nonetheless not straightforwardly borrowed into Finnish. Rather the internal characteristics of Finnish, that is, the use of partitive case and the use of the singular form after numerals, contribute to the rise of {N,NUM,OBL,SG}.

11.3.6.3 Genetic drift

In Biology, *genetic drift* refers to a mechanism of evolutionary change involving a *random* chance of genetic transmission. According to Darwinian theory, the chances of survival and reproduction among individuals in the population can be different due to *natural selection* (cf. §11.3.6.4). However, if the chance of survival and reproduction is not different among individuals in the population, the chance that different genes can be transmitted to the next generations is accordingly equal. The evolutionary change in a population therefore depends on the numbers of offspring of individuals in each generation. Suppose in a population there are two variants of genes; that is, gene A and gene B. In the 1st generation, the ratio between gene A (non-mutant gene) and gene B (mutant gene) is 80: 20. However, it seems unlikely that the parental generation of gene A and gene B will produce exactly the same numbers of offspring. Suppose further, the parental generation of gene B produces more offspring than that of gene A. In the 2nd generation, the ratio may change to 60: 40. If in the 3rd generation, the parental generation of gene A produces more offspring than that of gene B, the ratio between

gene A and gene B may change to 65: 35, for example. Over a number of generations, the population may end up with the dominance of gene A over gene B, or the equal frequency of the two genes, or the dominance of gene B over gene A. In the last case, we could say that an evolutionary change apparently takes place in the population. Normally, the chance that the mutant genes would replace the non-mutant gene would be very slim. However, due to genetic drift, a chance like that can happen, especially in the case in which the size of the population is very small.

We can see that genetic drift involves variation and fluctuating change. In language, variation is a common phenomenon. Like biological variation, linguistic variation is a sign of historical change. It is generally acknowledged that there are two types of variation. One is *free variation* and the other is *complementary distribution*. Free variation refers to a situation where more than one linguistic form is allowed to be used alternately in the same context. For example, English *what* can be pronounced [wa^h] or [wat] without a change in meaning. Complementary distribution refers to a situation where two different but related linguistic forms are used in different contexts. For example, in Old English there are at least two forms of CNNC_{NSG}. They are {N,NUM,NSG} for small numerals and {N,NUM,OBL,NSG} for higher numerals. In this case, the two constructions will not be in competition and can be present together in the language because each has its own distribution. However, the two variants may converge into one form as happened in the history of English when the higher numerals became treated as adjectives instead of nouns. The change from {N,NUM,OBL,NSG} to {N,NUM,NSG} was *directional*; that is, the use of {N,NUM,NSG} was extended to higher numerals only and not the opposite way.

11.3.6.4 Natural selection

In biology, *natural selection* is another major evolutionary mechanism. It refers to the natural process whereby some organisms having adaptive characteristics for their environment will have more chance to survive and reproduce their offspring. In contrast, those which do not have adaptive characteristics will have a smaller chance to survive and will tend to become extinct subsequently. Over generations, the organisms with

adaptive characteristics will spread over the population. Natural selection is an important mechanism of evolutionary change when the members of the population vary in adaptive characteristics. In a population, if all members have the same adaptive characteristics, natural selection will play no role.

In language, if we regard linguistic units (e.g. structural types of CNNCs) as organisms, language users as well as the linguistic systems as a whole⁴⁸ can be regarded as an environment in which those linguistic units are based⁴⁹. That is, the stability of linguistic forms depends on both. Linguistic forms which are favoured by language users and linguistic systems of the time and place are supposed to be selected, and those which are less adaptive will be vanishingly rare and will disappear. As with organisms, natural selection will play an important role only when there is variation in language. A question may arise: how can we say that a feature is adaptive? The adaptive linguistic form is the feature which is beneficial to the language users in terms of communicative efficiency (i.e. showing this adaptive feature: economy and expressivity) or learnability. At the same time, the linguistic form must be compatible with other linguistic systems.

The choice of a linguistic form which does not involve communicative efficiency should not be referred to as natural selection. For example, suppose that in English there were two ways to say 'thank you' e.g. *ta* and *thanks*. Suppose further that the young generation just happened to choose to say *ta*. Then *ta* would spread even though *thanks* is a word with the same efficiency. In this case, we can see that *ta* does not show more adaptive characteristics than 'thanks', so no natural selection is involved. In biology, this may be referred to as *neutral selection* (cf. Kimura 1983), or drift. Another noteworthy point is that some linguistic structures may show a tendency to be changed in the young generation (e.g. a minor sound change), but due to the prescriptive grammar in schools, other linguistic structures then remain unchanged. This is obviously unnatural and accordingly should not be referred to as *natural selection*. This phenomenon corresponds to biological *artificial selection* (i.e. the breeds are selected by humans).

⁴⁸ The reader is referred to a classic quotation "un langue est un système où tout se tient" [a language is a system where everything holds everything else] (translated by Jim Hurford, personal communication).

⁴⁹ Comparable to a biological ecosystem in which the organisms of the same species are also regarded as a part of the environment of that species.

In terms of CNNCs, natural selection plays an important role both at the global level and local level. At the global level, according to the frequency of structural patterns of $CNNC_{SG}$ and $CNNC_{NSG}$ reported in Chapter 6 and Chapter 7, the {N,NUM} construction is by far the most frequent type present in dominant and non-dominant modes, outnumbering other types in both $CNNC_{SG}$ and $CNNC_{NSG}$. This can possibly be explained in terms of natural selection. The {N,NUM} construction favours communicative efficiency; that is, the structure is expressive (i.e. giving all the required information) and economical (i.e. there are only the necessary elements; no redundant elements). At the local level, there are several cases where natural selection is a prominent. For example, the numeral classifiers in some languages are getting lost, perhaps because the speakers do not find them necessary. In pidgins and creoles, the number markers are mostly dropped except when the numerals are absent. These phenomena reflect the situation that these elements are non-adaptive for speakers of some languages.

In conclusion, the four mechanisms (i.e. mutation, gene flow, genetic drift and natural selection) all play important roles in evolutionary change in biological organisms and language in a combinatory way. It is difficult to determine which mechanism plays the most important role in evolutionary change. In biology, scientists used to agree that natural selection was the key evolutionary mechanism. However, currently some scientists are convinced that *neutral theory* or *genetic drift* plays the most important role (Kimura 1983; Caldwell et al. 2006)⁵⁰. In language, it is hard to determine whether natural selection is the major mechanism of language change since there are a number of changes in language which cannot be attributed to natural selection. We have looked into a number of changes that happened in the history of CNNCs, and it can be observed that some changes happened as a series of changes although the first change may be attributed to natural selection. For example, a change in the phonological system may

⁵⁰ It is noted in Caldwell et al. (2006) that “It might seem like everywhere we look, we see evidence of natural selection: organisms seem to be pretty well adapted to their environments. But the neutral theory of molecular evolution suggests that most of the genetic variation in populations is the result of mutation and genetic drift and not selection”. For further details, see <http://evolution.berkeley.edu/evosite/evo101/IIIE5b/Neutraltheory.shtml>

lead to a change in the morphological system as happened in Celtic languages; and the loss of gender leads to the loss of number marking in some Niger-Congo languages. In any event, we can see that although language is a cultural artefact, it looks more similar to biological organisms rather than to other cultural artefacts. In other cultural artefacts, natural selection may play a less important role. On the contrary, *neutral selection* as well as *artificial selection* may play a more important role here (e.g. old-fashioned clothes today will perhaps become popular in the future (see Bentley, Hahn and Shennan (2004), for instance, for more discussion about random drift in cultural change).

11.3.7 Adaptation, vestigial structures, maladaptation, and exaptation

As mentioned earlier, in biology, the term *adaptation* is generally referred to as an organism's advantageous characteristic that helps the organism to survive and reproduce in its natural habitat. For example, many lizards can change skin colour in response to the environment to protect themselves from predators. The ability to change skin colour is an adaptation. Yet, some present features of an organism are hard to see as having evolved for adaptive purposes. These features were once probably adaptations but later when their environment changed, these features become functionless. Such features are referred to as *vestigial structures*. Some vestigial structures may be harmful to the possessor. Such vestigial structures are known as *maladaptations*. Another relevant term is *exaptation*. As already introduced in §11.2.2.2, this term refers to the modification of the original function of a feature.

In this respect, language is not different from biological organisms. Those italicised terms can be borrowed to explain the nature of language change as evidenced in CNNCs. The CNNCs themselves are *adaptations* in reaction to the need for verbally expressing the exact number of things when the society developed beyond a primitive material stage. In addition, some adaptations can be observed from the grammatical features relevant to quantification in general. An example is the use of reduplication to distinguish number. The types of reduplication in some languages such as in Native American languages may be used so often that they have developed into grammatical markers (Mithun 1988). Another instance of adaptation with regard to number can be

found in the use of plural tone in some languages such as Ngiti, a Nilo-Saharan language (Dryer 2005a: 138).

However, most of the extra elements in CNNCs cannot be said to be adaptations, but rather exaptations. These features were not originally used in relation to quantification. For example, numeral classifiers were originally nouns representing physical things. When the language requires a word to unitize collective-like reference (analogous to English *police*, *cattle*), some nouns which have the potential to indicate units of things are therefore employed. So, such a reanalysis of nouns is referred to as grammaticalization in linguistics and corresponds to exaptation in biology. Another instance of exaptation can be observed in the grammaticalization of plural markers. However, not every instance of grammaticalization is an exaptation. For example, the development of dual and trial markers in place of numerals is not an exaptation. This process happens just because of sound erosion (due to frequent use). These items are then reanalysed as grammatical numbers. In this case the grammaticalization process should be referred to as an adaptation because the erosion of sound results from ease of articulation—it is a kind of communicative efficiency.

The use of number markers as either adaptations or exaptations in the first place is rather transparent, namely, the motivation for their use may have been instigated by the first innovating speakers. However, when there was a change in the linguistic system, namely the emergence of numerals, the number markers became redundant and *vestigial* since number was indicated by numerals already. In many Niger-Congo languages, the grammatical morphemes denoting person-gender-number (PGN) may instantiate vestigial structures in language because they are functionless or marginally functional. At the beginning the use was functional since the morpheme was used to clarify the number and person. The loss of PGN agreement as happened in many Niger-Congo languages reflects the situation that the PGN morphemes are maladaptive and so the language system has tried to get rid of them. This is because it is a burden on speakers to memorise them for no benefit at all.

To sum up, the biological notions of *adaptation*, *vestigial structure*, *maladaptation* and *exaptation* help to understand how and why biological features evolve. These terms can be applied well to language so as to understand the various origins of present characteristics of modern languages.

11.3.8 Convergent evolution

In biology, the term *convergent evolution* refers to the process in which different species evolve similar characteristics because of similar selective pressures. Such characteristics are referred to as *analogous*. For example, whales, which were terrestrial animals, have evolved flippers for swimming in reaction to the marine environmental condition. The flippers function like the fins of fish. However, the flippers and fins have different origins.

Turning to language, convergent evolution happens normally in language. A large number of features in genetically unrelated languages may become similar if they encounter the similar linguistic conditions. Convergent evolution can often be found in CNNCs. For example, the {N,NUM,CLF} construction found in Chinese, African languages, and Native American languages does not obviously develop from a common origin but rather these languages have encountered a similar selective pressure (i.e. semantic clarification).

However, in language, analogous features do not necessarily involve a similar selective pressure. For example, the {N,NUM} constructions in several languages have different origins. In CNNCs, there are several cases instantiating convergent evolution. For example, {N,NUM} in Creoles, Celtic languages, and some Uralic languages did not develop from the same selective pressures.

In sum, the idea of convergent evolution is important in language change. That is to say, it helps us not to jump mistakenly to conclusions about genetic relationship or language contact when two features are similar (such as {N,NUM,CLF} in Asian and Native American languages). Rather, the similarity of the features is possibly due to similar selective pressures. In addition, similar forms can emerge by accident just like the {N,NUM}, in which case the idea of convergent evolution is not applied.

11.3. 9 Uniformitarianism

Uniformitarianism is an assumption employed in studying the past in historical sciences. The concept was originally used in geology and was later adapted in evolutionary biology and linguistics. In geology, the concept was first introduced by the geologists James Hutton and was modified later by Charles Lyell around the early 19th century, referring to the assumption that the geological laws and processes (including a rate of change) in the past history of the earth were essentially uniform with the ongoing processes in the present. For example, the formation of mountains and the uplift of land caused in the past by volcanic action and earthquakes respectively can be seen in the present day (Charlesworth and Charlesworth 2003: 43). This assumption leads to the analogical method of inferring or reconstructing past events from the laws or processes observed in the present. The assumption also suggests that the earth has changed gradually through cumulative action of natural processes over a long period of time rather than abruptly by catastrophic events. The concept of uniformitarianism was later adapted in Darwin's evolutionary theory. It is assumed in evolutionary biology that the evolutionary processes take place slowly at all times at a steady rate; the uniformitarian principle in this sense is known as *gradualism* (Bak and Paczuski 1996: 1). Fossilized transitional forms and organic variation may support the facts of gradual change. In addition, based on the steady rate of change, the long term trends can be inferred from the facts or data observed over a short period in the present. However, the hypothesis about uniformitarianism remains controversial, since some processes are somewhat particular in explaining geological and biological phenomena (Caldwell et al. 2006). Many scientists tend to believe that although uniformitarianism may dominate natural change, catastrophic events such as violent earthquakes or huge asteroid impacts can happen intermittently, having a global effect (Bak and Paczuski 1996: 2-6).

The uniformitarian principle came to be used in diachronic linguistics as well. Basically, the principle assumes that the general nature of human languages of the past was not so different from that of the present (cf. Newmeyer 2002: 359; Croft 2003: 233). This principle can be best applied to the distant past of human language only as far back as comparative reconstruction allows, say 5000-6000 years ago. In comparative

reconstruction, the uniformitarian principle is used as a reconstruction method constraining the hypotheses about the structure of the past language. That is, the characteristics of the ancient language hypothesised must be uniform with those of contemporary languages (Ringe, Warnow and Taylor 2002: 60-61). To illustrate, according to *diachronic typology*, the same typological generalizations observed in modern languages should apply to ancient languages (Croft 2003: 233). In the case just mentioned above, the uniformitarian principle is interpreted as corresponding to *language universals—the uniformity of laws* governing human language across time and space.

However, in some places the uniformitarian assumption is interpreted as *the uniformity of state*. In geology, the uniformity of state means that the state of the earth in the present is not different from that of the past. The notion *uniformitarianism* in this sense in geology is no longer accepted, though. In linguistics, the uniformitarian assumption in this sense suggests that human language has not changed since the earlier stages, say since fifty or one hundred thousand years ago. Yet, some scholars have presented evidence that seems to suggest that the earliest human languages might have looked different from the human languages spoken today, and therefore, the uniformitarianism assumption does not seem to be totally correct. According to Gil (2006: 92-93), some grammatical processes are argued to be unidirectional, such as grammaticalization whereby lexical items, typically nouns and verbs, tend to develop into grammatical items such as prepositions, rather than grammatical items developing into lexical items (see more details in Heine and Kuteva 2002b, 2007). Another unidirectional change is that languages are more likely to exhibit a typological shift from OV (object precedes verb) to VO (verb precedes object) than vice versa (see more details in Givón 1979, for instance). Newmeyer (2002) reviews some works to support the idea that grammar depends on culture. Because culture in the past is different from the present, it is possible that languages in the distant past might have been different from the present to some extent. For example, in Perkins (1992 in Newmeyer 2002), the grammatical items expressing deixis (e.g. English ‘this’ ‘that’) in a materially complex society are likely to be fewer than in a materially primitive society. If indeed all of these

hypotheses are correct, early language might have had very few grammatical items, OV word order, and several grammatical forms for deixis, for instance.

Turning to CNNCs, the data support the conclusion above; that is, human language shows uniformity of laws but not of state. With regard to the uniformity of laws, the typological generalization that “languages with {N,NUM,NSG} tend to have genders and vice versa” not only applies to contemporary languages but also to ancient languages like Sanskrit or Latin. Nevertheless, the uniformity of laws in language with regard to diachronic typology is subject to change. English, for instance, exhibits the construction {N,NUM,NSG} but the genders of nouns have been lost. This suggests that the uniformity of laws in language is not so rigid as in geology and biology.

As for the uniformity of state, CNNCs were obviously not found in early language until the birth of numerals, some of which developed from existing lexical items through the process referred to as *numeralization*. Also, the birth of numerals exhibits unidirectional change; that is, from the small numerals towards the higher ones over a historical period. No language is reported to have high numerals without small numerals. The small numerals in turn developed into number markers, giving rise to complex number categories in later languages. All of these are clearly not present in early language. In addition, according to the diachronic development of $CNNC_{NSG}$ as already illustrated in the previous chapters, it can be argued further that in early language the proportion of *set nouns* was significantly higher than that of other types of nouns, such as *singular object nouns*, and *sort nouns*. The CNNC in the initial hypothetical stage is {N,NUM}. According to Rijkoff (2003), nouns in this construction are set nouns. As for singular object nouns, it is hypothesised that this type of noun emerged later, after the nouns had been used with plural markers so that the plural nouns must be accompanied by plural markers. The process led to the emergence of singular object nouns which are employed in the {N,NUM,NSG} construction. Regarding *sort nouns* which are employed in the {N,NUM,CLF} construction, based on the diachronic development in Chapter 10, they are likely to appear later than the set nouns and singular object nouns.

Finally, the current study also suggests that language may exhibit an unsteady rate of change. For example, the grammatical categories linked to quantification (i.e. number systems and classifier systems), if they had already existed in any form in early language, might have flourished faster upon the arrival of numerals.

To sum up, it is concluded that with regard to uniformitarianism, biology and language are quite similar in general. Biological change has held true to a uniformitarian scenario with regard to the uniformity of laws. In this case, language evolution shows the same scenario, but with less rigidity. Uniformitarianism of state seems to be incorrect both in biology and language.

11.4 Conclusion

This chapter has discussed the application of an evolutionary biological approach to language change by comparing the two evolutionary systems in regard to some major issues. The comparison has shown that evolutionary biology can be used as a *unifying* approach to understand the nature of human language as a unique cultural artefact.

The nature of language based on the discussion so far can be surmised as follows. Language is an evolutionary system as it can be replicated with modification. The origins of linguistic structures are not straightforward. Some may emerge as simple forms, but others may emerge as complex forms. Language can develop either towards simplicity or complexity, but the most important point is that it develops to maintain communicative efficiency and effectiveness at a given time. Language generally changes gradually, but a sudden change is expected to happen intermittently.

When the speech communities of a language are split up, over a period of time, new dialects of the language may appear. The new dialects can change into different languages due to geographic isolation. However, the different characteristics of the two languages may be recombined again if the two languages come into contact. So, theoretically, reproductive isolation will never happen in language.

The major mechanisms driving language evolution are functional motivation (comparable to *natural selection*) and language contact (comparable to *gene flow*). Free variation (comparable to random mutation as well as neutral selection) seems to be an important mechanism as well, but it is not clear from the current research to what extent it is involved in language change. It seems likely that drift or fluctuating change (corresponding to genetic drift) in language may be quite insignificant. In any event, these mechanisms work in a combinatory way.

Over the history of human language, grammatical features have developed. Current grammatical features may have different origins. Some are adaptations and the others exaptations. Once those features are no longer functional, language will turn them into vestigial structures; or get rid of them if they are maladaptations. Some grammatical features in genetically unrelated languages may be identical if the languages encounter

the same selective pressures. This phenomenon can be seen as convergent evolution. The final aspect is uniformitarianism. With respect to the uniformity of laws, the uniformitarian principle is feasible. Uniformitarianism with respect to uniformity of state is questionable.

Finally, the comparison suggests that language is strikingly similar to biological organisms in general. A biological evolutionary analogy sounds workable for language change. The differences or mismatches mentioned sporadically above involve a degree of flexibility in some circumstances, such as gene flow/language contact and the uniformity of laws. Despite some mismatches, it seems likely that language is perhaps more similar to biological organisms than to other cultural artefacts with regard to evolution. An account for the uniqueness of the nature of language is that language is a combinatory product of *biological, psychological and cultural phenomena*.

12 Conclusions

The previous chapters have shown that there are several structural patterns of cardinal numeral-noun constructions (CNNCs) in the world's languages. This phenomenon is interesting because it appears to violate the principles of economy and distinctness in language—the principles that seem to be key selection pressures for language evolution. In other words, change that happens in language systems should be explainable in the light of these two principles. However, due to the diversity and complexity of most structural types of CNNCs, the immediate question then arises as to how the diversity and complexity happen, despite the fact that simple structures (i.e. {N,NUM} like English *three sheep* seem very practical.

Historical studies are always required when we want to understand the present. Responding to the question above, the thesis assumes that viewing CNNCs from a diachronic perspective would aid our understanding of the complexity and diversity of structural patterns of CNNCs existing in modern languages. Even though counting expressions seem to be rather basic in human daily life, the whole evolutionary scenario of CNNCs had yet to be drawn. Therefore, the thesis has set out an ultimate aim of developing an evolutionary picture of CNNCs with the belief that it would contribute to the understanding of the diversity and complexity of CNNCs and of human language in general.

To achieve this goal, the thesis has explored the structural patterns of CNNCs of some 240 languages across the globe through reference grammars. The historical origins of these structural types were then investigated, using evidence from old written records along with theoretical approaches. Finally, with cross-linguistic comparison integrated with diachronic approaches, hypothesized evolutionary trajectories of CNNCs were postulated. In this chapter, the findings of the present study are summarized along with some explanations. The chapter ends with certain concluding remarks.

12.1 Summary of findings

12.1.1 Typological issues of CNNCs

Regarding the possible *structural types* of CNNCs attested in living languages, the findings as described in Chapter 6 illustrate that there is a great diversity of structural types of both $CNNC_{SG}$ and $CNNC_{NSG}$. As for $CNNC_{SG}$, in addition to the structural types of $\{N,NUM\}$ and $\{N,NUM,CLF\}$ which are very common, there are other structural types which are relatively rare, namely, $\{N,NUM,SG\}$ (e.g. Persian), $\{N,CLF\}$ (e.g. Vietnamese), $\{N,Unit\}$ (e.g. Somali), $\{NUM,CLF\}$ (e.g. Nepali), $\{N,NUM,CLF,SG\}$ (e.g. Ejagham), $\{N,NUM,NPI\}$ (e.g. Imonda), $\{N,SG\}$ (e.g. Mohawk), and $\{N,NUM,OBL,SG\}$ (e.g. Berber (Ayt Seghrouchen Middle Atlas)). The atypical type is $\{N,NUM,INV\}$, where a noun must have inherent number as in Jemez. There are two languages, namely Pirahã or Wari¹, reported as lacking a numeral *one* proper and hence completely lacking $CNNC_{SG}$; these languages use a word with a numerical interpretation meaning ‘small’ and ‘alone’ respectively.

As for $CNNC_{NSG}$, in addition to $\{N,NUM\}$, $\{N,NUM,NSG\}$, $\{N,NUM,CLF\}$, $\{N,NUM,CLF,NSG\}$, $\{N,NUM,OBL,NSG\}$, and $\{N,NUM,OBL,SG\}$ which have already been mentioned in earlier research, there are some other structural types found, though they are rather rare. These include $\{N,NUM,RMS\}$ (e.g. Somali), $\{N,NUM,ACC,SG\}$ (e.g. Arabic), $\{NUM,CLF\}$ (e.g. Kuna), $\{N,NUM,DPM\}$ (e.g. Degema), $\{N,NUM,SG\}$ (e.g. Berber (Ayt Seghrouchen Middle Atlas)), $\{N,NUM,NUMPCL\}$ (e.g. Maori) $\{N,NUM,PNUM,NSG\}$ (e.g. Tuvaluan), $\{N,NUM,NPL\}$ (e.g. Imonda). Like $CNNC_{SG}$, there exist languages with no numeral and hence no $CNNC_{NSG}$ at all (e.g. Pirahã).

Before moving on to the historical origins of these structural types, the thesis performs two typological tasks. The first is typological classification—that is, languages are classified into single types. This task not only illustrates the classification, frequencies and distribution of CNNCs, but the findings are also a requirement for the second typological task. The second task has to do with (implicational) universals which

are typological generalizations. Making typological generalizations requires knowing language types, their frequencies, and distribution.

According to the findings of the typological classification, the languages may be classified into 6 types for $CNNC_{SG}$ and 11 types for $CNNC_{NSG}$ based on the primary structural types used in the language.

With respect to $CNNC_{SG}$, the world's languages are roughly divided into 2 main types, namely, $\{N,NUM\}$ and $\{N,NUM,CLF\}$ with very few other possibilities. $\{N,NUM\}$ is by far the most common, greatly outnumbering all other types combined. This simply suggests that the structural pattern of $\{N,NUM\}$ is unmarked in most languages of the world. The other major type is $\{N,NUM,CLF\}$, consistent with earlier research, and this type is fairly common in a few particular areas, notably the Pacific Rim and eastern India. However, some languages, although relatively rare, do not fall into the two groups mentioned above. For example, several instances in Niger-Congo languages use $\{N,NUM,SG\}$ as a primary method. A few languages show the co-existence of types with none primary. Several languages in contact with numeral classifier languages show a mixture of $\{N,NUM\}$ and $\{N,NUM,CLF\}$ as in Persian. The last type is represented by the languages where the (native) numeral *one* proper is absent. This finding implies that humans prefer to use simple structures to express $CNNC_{SG}$.

Regarding $CNNC_{NSG}$, the languages of the world can be basically classified into 4 major types, namely $\{N,NUM\}$, $\{N,NUM,NSG\}$, $\{N,NUM,CLF\}$ and a mixture of $\{N,NUM\}+\{N,NUM,NSG\}$ with some other possibilities. $\{N,NUM\}$ is the most common. This type spreads throughout most major regions of the world except Western Europe, Southeast and (parts of) East Asia. The second most common is $\{N,NUM,NSG\}$. This type is fairly common in several major regions of the world but absent in Southeast Asia. The third most common is $\{N,NUM,CLF\}$. The distribution of this type is not different from what is already described in $CNNC_{SG}$. In addition to the 4 major types, there are other possibilities used as primary or non-primary observed in a few instances, for example, $\{N,NUM,OBL,NSG\}$, found mostly in Europe and

{N,NSG} (where NSG = DU and TRI), observed in a few languages of Australia where numerals are limited to low-valued numbers. Finally, many languages show a mixture of types; notable is a mixture of {N,NUM} and {N,NUM,NSG} with neither primary. The mixture may suggest a change happening in the languages.

12.1.2 Historical origins of CNNCs

The next step is to investigate the historical origins of structural types of CNNCs, using old written records together with theoretical considerations. The findings as shown in Chapters 7 and 8 can be summarized as below. Some patterns of diachronic change proposed below may be common, while others probably are not.

CNNC_{SG}

{N,NUM}	<	N+ word with a numerical interpretation (e.g. Sanskrit)
{N,NUM}	<	{N,NUM,CLF} (e.g. Chinese)
{N,NUM,SG}	<	{N,NUM} (e.g. Pame)
{N,NUM,SG}	<	{N,SG} (e.g. Lunda)
{N,NUM,CLF}	<	{N,NUM} (e.g. Chinese)
{N,NUM,CLF}	<	{N,NUM,SG} (e.g. Kana)
{N,SG}	<	{N,NUM} (e.g. Burushaski)

CNNC_{NSG}

{N,NUM}	<	N+word with a numerical interpretation} (e.g. Wari ')
{N,NUM}	<	{N,NUM,NSG} (e.g. Irish)
{N,NUM}	<	{N,NUM,CLF} (e.g. Nivkh)
{N,NUM,NSG}	<	{N,NUM} (e.g. Hawaiian Creole)
{N,NUM,NSG}	<	{N,NUM,OBL,NSG} (e.g. Norwegian)
{N,NUM,CLF}	<	{N,NUM,NSG} (e.g. Bengali)
{N,NUM,CLF,NSG}	<	{N,NUM,NSG} (e.g. Malto)
{N,NUM,OBL,SG}	<	{N,NUM,NSG} (e.g. Russian)
{N,NUM,OBL,SG}	<	{N,NUM} (e.g. Finnish)

- {N,NUM,OBL,NSG} < {N,NUM} (e.g. Welsh)
- {N,NSG} < {N, word with a numerical interpretation}
- {N,NSG} < {N,3PL} (e.g. Gaagudju)
- {N,NSG}(NSG=DU,TRI) < {N,NUM} (e.g. Kayardild)

12.1.3 Evolutionary scenario for CNNCs: a possibility

As presented in Chapter 9, based on the assumption that the various structural patterns of CNNCs discovered can be considered in the light of a hypothetical evolutionary ladder, these historical pathways just described above are combined together to postulate larger subsystems in several layers, even though most of them are not genetically related. We then end up with a *possible* general evolutionary trajectory of CNNC_{SG} and CNNC_{NSG} as represented in Figures 12.1 and 12.2 respectively.

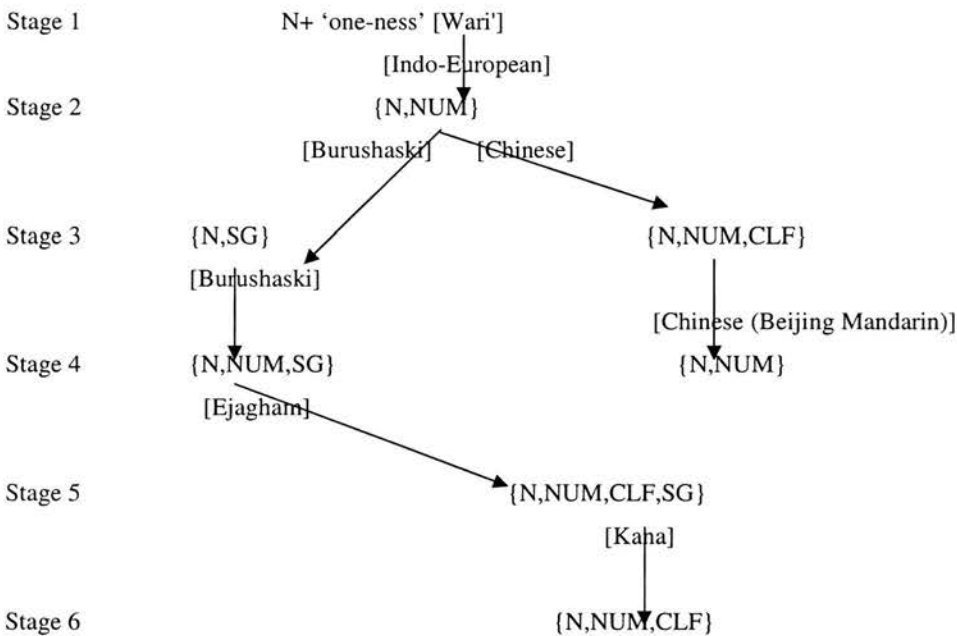


Fig. 12.1 A Possible Evolutionary Trajectory of CNNC_{SG}

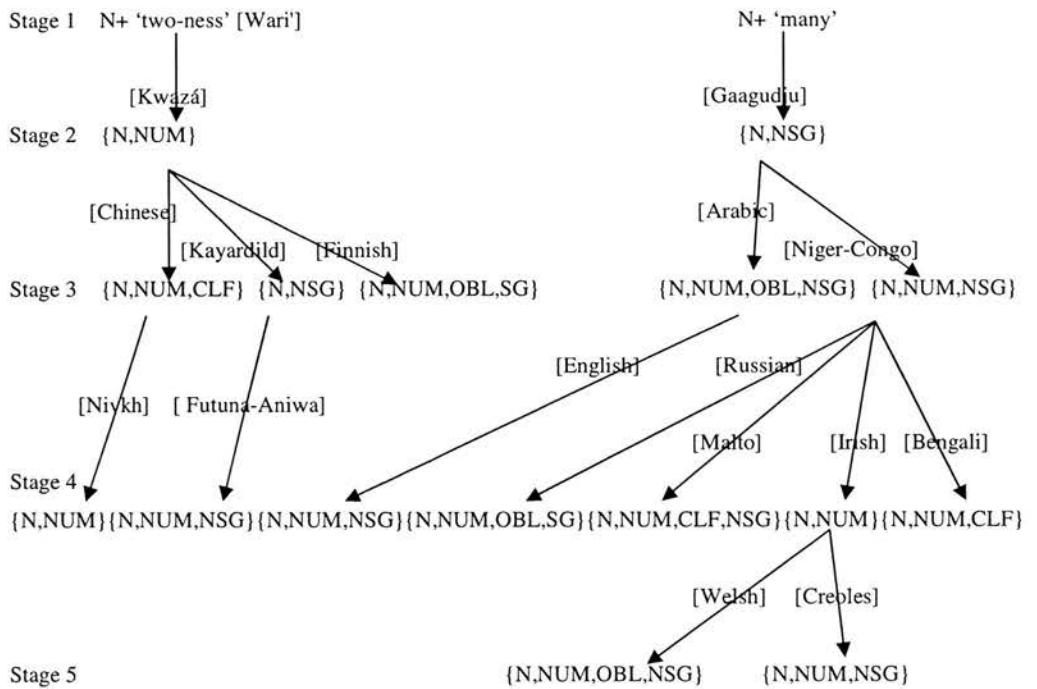


Fig. 12.2 A Possible Evolutionary Trajectory of CNNC_{NSG}

From the evolutionary trajectories above, it can be noticed that the development of CNNCs is *bi-directional* or *reversible* in terms of structural complexity. The constructions may develop toward greater complexity initially but they can develop toward greater complexity or greater simplicity over the course of evolution. Another major characteristic of CNNC_{NSG} is that many patterns of CNNC_{NSG} are *unpredictable* or *idiosyncratic*. That is, many patterns of change take place in a particular linguistic context. For example, {N,NUM,OBL,SG} emerged in Russian because the dual marker was reinterpreted as genitive singular. This grammatical reinterpretation seems unlikely to happen as a common occurrence.

12.1.4 Explanations for the complexity and diversity of CNNCs

All the changes in the evolutionary scenario result in the complexity and diversity of CNNCs in the modern languages. As discussed in Chapter 10, these changes are triggered by various factors which can be categorized into three groups, namely a quantifying function, a non-quantifying function, and a mixture of both. Most of these factors in fact can be explained in the light of the principles of economy and distinctness.

The factors concerning the quantifying function involve the rise of numerals, the presence of collective nouns and general nouns, and the special meaning. The rise of numerals is a direct factor in the emergence of CNNCs. There are several structural patterns that emerged by this motivation. Examples are {N,NUM}) that developed from a noun plus a word with a numerical interpretation, and {N,NUM,NSG} where numerals were presumably attached with non-singular markers before the rise of numerals. The presence of collective nouns (e.g. English *cattle*) and general nouns (e.g. Chinese *xìn* ‘a letter’ or ‘letters’) is included in this kind of motivation. Following the general view, both kinds of nouns cannot be in a direct construction with numerals, rather the individualizer is required to make the enumeration possible. The individualizer can be a singulative marker or a numeral classifier. This leads to the complexity of the construction. The structural patterns that emerged because of these motivations are {N,NUM,CLF} and {N,NUM,SG}. Besides, the presence of nouns that can be counted in more than one way, such as the use of a numeral classifier in some nouns denoting plants in Ejagham, also leads to the rise of {N,NUM,CLF}. Moreover, the special meaning may help to maintain the variety of structural patterns in the language such as the slight difference in meaning between {N,DU} and {N,NUM} in Ngiyambaa (Pama-Nyungan, Australia). The changes just described above are clearly attributed to the principle of distinctness.

The diversity in CNNCs is also attributed to some other factors which are not relevant to a quantifying function. Rather, it is attributed to a change in other areas of grammar. The loss of cases and noun classes with which grammatical number is fused

may trigger a change in CNNCs. This happened in some of the Goidelic and Niger-Congo languages where {N,NUM,NSG} changed into {N,NUM} because of the loss of cases and noun classes respectively. Also, the fusion of a numeral classifier and a noun as illustrated by Chinese results in a change in CNNCs from {N,NUM,CLF} to {N,NUM.CLF}, which superficially looks like {N,NUM}. The change in grammatical categories of numerals can change the structural type of CNNCs as happened in English where the pattern of {N,NUM,NSG} (e.g. Modern English *four hundred winters*) is derived from the pattern of {N,NUM,OBL,NSG} (e.g. Old English *four hundred of winters*) because in Old English, the high round numerals such as *hundred* are treated as nouns but later as more adjectival in Modern English. A change such as this probably reflects a systemization in language. The changes illustrated above are attributable to the principle of economy in language. The reanalysis of the dual marker as the genitive singular, giving rise to the change from {N,NUM,NSG} to {N,NUM,OBL,SG} as happened in Russian is also another instance of the non-quantifying motivation. This change cannot be attributed to either the principle of economy or distinctness.

Moreover, the diversity of CNNCs may be caused by factors relevant to a mixture of quantifying and non-quantifying functions. For example, in some Arabic dialects, the rise of {N,NUM,NSG}(NSG =PL) for the numeral *two* is derived from {N,NSG} (NSG=DU). The initial cause would be the loss of concord or agreement (i.e. a non-quantifying factor) leading to the loss of the dual in the context of the numeral *two*. This is because the dual might be seen as redundant (i.e. a quantifying factor). Later, the use of plural markers was extended to dual nouns by analogy with other plural nouns (i.e. a non-quantifying factor). The two changes are attributable to the principle of economy. Another instance can be seen in the grammatical change in the dual marker in Futuna-Aniwa. In this language, {N,DU} changed into {N,NUM,DU} when the dual marker assumed the new grammatical function of an article (i.e. a non-quantifying factor), and the numeral *two* was then required again. This change is attributable to the principle of distinctness.

Overall, the results suggest that the principles of economy and distinctness still play major roles in the complexity and diversity of CNNCs. It can also be further implied that the motivations for the changes are functional. However, some changes to certain CNNCs seem irrelevant to the two principles, as happened in {N,NUM,OBL,SG} in Finnish and {N,NUM,OBL,SG} in Russian which are results of language contact and misinterpretation respectively.

12.2 Concluding remarks

The thesis offers clear evidence for general characteristics of human language which look similar to the nature of biological organisms in general, as discussed in Chapter 11. For example, human language can develop towards either simplicity or complexity so as not to reduce communicative efficiency and effectiveness at a given time. Language changes gradually but an abrupt change may be observable occasionally if there is a punctuating factor. The major mechanisms driving language to evolve mostly involve functional motivation and language contact. Over the history of human language, grammatical features have developed. Some of the grammatical features are adaptations and some are exaptations. From among these features, some may change to vestigial structures if they are no longer functional; or may even be removed from the language due to their maladaptedness. Some grammatical features in genetically unrelated languages may look alike if the languages encounter the same selective pressures. This phenomenon is comparable to convergent evolution in biology. The final aspect has to do with uniformitarianism. The uniformitarian principle seems likely in terms of uniformity of laws. Nevertheless, due to language contact, some current linguistic patterns appear to violate these laws. Uniformitarianism with respect to uniformity of state is doubtful, especially for the linguistic features concerning cultural development. Finally, this comparison also implies that languages are strikingly similar to biological organisms in general. The biological evolutionary analogy seems workable for a study of language change.

In addition, it should be emphasised that the conjecture on the evolutionary trajectories of CNNCs was based on a limited number of languages where data were sufficient for the possibility of diachronic analysis. This means that there may be some other possibilities in addition to those illustrated. For example, as already mentioned in § 2.2.1, a numeral classifier language may change into a plural marking language as evidenced by Chamorro (Austronesian, Guam), a language which underwent a typological change after contact with Spanish and English (Sanches and Slobin 1973:

12). Further, {N,NUM,CLF} may change into {N,NUM,NSG} but this is not found in my data. Additionally, there may be some other contributory factors that are also not present in the current data. For example, {N,NUM} may derive from {N,NUM,CLF} because of some other reasons besides fusion (e.g. the omission and the decline of numeral classifiers in the language) but clear data on this are not available. Therefore, the real evolutionary trajectories are probably messier and the contributory factors may be more diverse.

Regarding methodology, the hypothetical evolutionary ladders as illustrated in Figures 10.1 and 10.2 were drawn by using cross-linguistic comparison integrated with a diachronic approach. This methodology assumes that the diversity of structural patterns present in languages today can be used as evidence for an evolutionary ladder. However, this does not necessarily mean that such a principle can be applied to every case of diversity in linguistic patterns. For example, cross-linguistically, there are three types of causative constructions in terms of morpho-syntactic properties, namely lexical, morphological, and periphrastic causatives. It may not always be the case that periphrastic causatives are the initial stage and then that morphological causatives and lexical causatives follow in any possible order. There has been research (see, for example, Haiman 1985) suggesting that the three causative constructions may be used differently due to semantic differences if the three types are found in the same language. For example, periphrastic causatives (e.g. *make something broken*) are less iconic than lexical ones (e.g. *break something*) if the two types are in the same language. Therefore, we need to be cautious about this methodology.

Finally, there are some topics on which future research should focus. Firstly, as diachronic paths in the evolutionary trajectories have very few languages; further research should reveal whether each diachronic path has other members (other languages where this pattern of change occurs). The result would illustrate how common each diachronic path actually is. It seems likely that some diachronic patterns are regular and predictable, but some are not. For example, the construction consisting of a noun plus a word with numerical interpretation is likely to change to {N,NUM} through the process of numeralization. On the other hand, {N,NUM,CLF} may not need to come from

{N,NUM} or {N,NUM,NSG}. There are probably also some other possible diachronic routes and contributory factors but clear evidence is not available. So, further research is needed for looking into other possibilities. For example, based on this thesis, we do have {N,NUM.CLF} which superficially looks like {N,NUM}, but we do not have {N,NUM} which is derived from {N,NUM,CLF}. It seems that this route may exist but there is no supporting evidence.

In the present study, there are some constructions for which the historical origins are not yet known, such as {N,NUM,UNIT} and {N,NUM,RMS}. So, the evolutionary trajectories would look more inclusive if the historical origins of these structural patterns were uncovered. In addition, there might be some constraints on language contact regarding CNNCs. There is no evidence suggesting that {N,NUM} developed due to language contact, for example. The development of {N,NUM} seems to be attributable to internal motivation only. Regarding extra elements, compared to number markers, numeral classifier systems are likely to be easier to spread to other languages. Moreover, one avenue for further research on numerical expressions would be to investigate related constructions cross-linguistically, such as ordinal numeral-noun constructions and approximative numeral-noun constructions, in terms of diversity and diachronic development. These likely areas of future research will contribute to the body of knowledge about the diversity and evolution of numerical expressions.

Appendix 1 Lists of Languages (by Area)⁵¹ and Structural Types of CNNCs

In Appendix 1, all the structural types of CNNCs in sample languages are presented in tabular form. For CNNC_{SG}, the reader is referred to Table A, and for CNNC_{NSG}, refer to Table B. The two tables also provide a complete listing of the languages in the sample and their geographical and genetic affiliation. Following the genealogical language list in WALS (2005: 584-644), the families in Table A and Table B are organized in the geographical fashion, starting from Africa (AFR), Eurasia (EUR), Southeast Asia & Oceania (SEA & OCE), Australia & New Guinea (AUS-NEW), North America (NAM) and ending in South America (SAM). Within each family, the genera and languages are organized alphabetically. As for creoles and pidgins, they are organised separately as an individual group in WALS. In the thesis, however, they are organised by area at the end of each region.

The structural types of CNNCs displayed in the two tables are either the basic types (represented by the symbol +) or subsidiary ones (represented by the symbol @). Note again that the structural types indicated in all languages are based only on the examples found in the sources. In some languages there probably exist some other types missing from the survey. The extensive examples from the language sample illustrating all structural types of CNNCs can be found in the grammar reference survey in Appendix 3.

It must be emphasised that the two tables in Appendix 1 shows *structural type*, not *language type* (see *Introduction in Chapter 6 for their definitions*). For a list of languages in relation to each *language type* established in Table 6.1 and Table 6.2, the reader is referred to Appendix 2.

⁵¹ For a list of languages organized alphabetically by the name, refer to Appendix 2.1.

Macro- Area	Family	Genus	Language	Country	{N, NUM}	{N, NUM, SG}	{N, NUM, CLF}	{N,SG}	Other
AFR	Khoisan	Central Khoisan	!Ani	Botswana		+			
1									
2	Niger- Congo	Bantoid	Ejagham	Nigeria		+			@ {N,NUM,CLF,SG}
3			Lunda	Democratic Republic of Congo		+			
4		Edoid	Degema	Nigeria		+			
5		Cross River	Kana	Nigeria			+		
6		Gur	Koromfe	Burkina Faso, Mali		+			
7		Ijoid	Ijo (Kolokuma)	Nigeria			+		
8		Kordofanian	Orig	Sudan		+			
9		Kru	Grebo	Liberia			+		
10		Northern Atlantic	Wolof	Gambia, Senegal		+			
11		Western Mande	Kpelle	Liberia		+			
12	Nilo- Saharan	Berta	Berta	Ethiopia, Sudan		+			
13		Fur	Fur	Sudan		+			
14		Moru-Ma'di	Ma'di	Sudan, Uganda		+			
15		Nilotic	Lango	Uganda		+			
16		Saharan	Kanuri	Chad,Niger, Nigeria,Sudan		+			
17		Songhay	Koyraboro Senni	Mali, Niger		+			
18	Afro- Asiatic	Berber	Berber (Ayt Seghrouchen Middle Atlas)	Morocco		+			+ {N,NUM,OBL,SG}
19		East Chadic	Lele	Chad		+			

Table A: Sample languages and their structural types of CNNC_{SG}

Macro -Area	Family	Genus	Language	Country				Type			
								{N, NUM}	{N, NUM, SG}	{N, NUM, CLF}	{N,SG}
20		Eastern Cushitic	Somali	Somalia	+						@ (N,UNIT)
21		Omotic	Maale	Ethiopia	+						
22		Semitic	Arabic (Modern Standard)	Saudi Arabia and adjacent countries in Middle East and North Africa	+						
23			Chaha	Ethiopia	+						
24			Hebrew	Israel	+						
25			Maltese	Malta	+						
26		West Chadic	Hausa	Niger, Nigeria	+						
27	Creoles and Pidgins		Angolar	São Tomé e Príncipe	+						
28			Nigerian Pidgin	Nigeria	+						
29			Nubi	Uganda	+						
EUR 30	Basque	Basque	Basque	France, Spain	+						
31	Indo-European	Albanian	Albanian	Serbia & Montenegro, Albania	+						
32		Baltic	Lithuanian	Lithuania	+						
33		Celtic	Irish	Ireland	+						
34			Welshwichi	UK (Wales)	+						
35			Gaelic (Scots)	UK (Scotland)	+						
36		Germanic	German	Austria, Germany, Switzerland	+						
37			English	United Kingdom	+						

Macro -Area	Family	Genus	Language	Country	{N, NUM}	{N, NUM, SG}	{N, NUM, CLF}	{N, NUM, CLF}	Other
38			Norwegian	Norway	+				
39	Indic		Bengali	India, Bangladesh			+		
40			Nepali	Nepal			+	@	{NUM,CLF}
41	Greek		Greek	Greece	+				
42	Iranian		Persian	Iran	+	@		@	
43	Romance		Catalan	Spain	+				
44			French	France	+				
45	Slavic		Russian	Russia	+				
46	Finnic		Estonian	Estonia	+				
47			Finnish	Finland	+				
48	Ugric		Hungarian	Hungary	+				
49	Samoyedic		Selkup	Russia	+				
50	Altaic		Mangghuer	China			+		
51			Mongol (Khamnigan)	China, Mongolia	+				
52		Tungusic	Manchu	China	+				
53			Udihe	Russia	+				
54		Turkic	Turkish	Turkey	+				
55	Yukaghir		Yukaghir (Kolyma)	Russia	+				
56			Yukaghir (Tundra)	Russia	+				
57	Yeniseian		Ket	Russia	+				
58	Chukotko-Kamchatkan		Alutor	Russia (Asia)		+			
59	Nivkh		Nivkh	Russia			+		
60	Ainu		Ainu	Japan	+		@		
61	Japanese		Japanese	Japan			+		

Table A: Sample languages and their structural types of CNNC_{SG}

Macro -Area	Family	Genus	Language	Country	Type		
					{N, NUM}	{N, NUM, SG}	{N, NUM, CLF}
62	Korean	Korean	Korean	Korea			
63	Northwest Caucasian	Northwest Caucasian	Abkhaz	Georgia	+		
64	Nakh-Daghestanian	Lezgic	Lezgian	Russia	+		
65	Kartvelian	Kartvelian	Georgian	Georgia	+		
66	Burushaski	Burushaski	Burushaski	Pakistan	+	@	@
67	Nahali	Nahali	Nahali	India	+		
68	Dravidian	Central Dravidian	Kolami	India	+		
69		Northern Dravidian	Malto	India			+
70	Austro-Asiatic (in Eurasia)	Munda	Bhumij	India	+		
71			Korku	India	+		
72			Mundari	India	+		
SEA&OCE	Sino-Tibetan	Bodic	Camling	Nepal			+
73							
74			Chantyal	Nepal			+
75			Tamang	Nepal	+		
76		Burmese-Lolo	Akha	China, Myanmar, Thailand			+
77		Chinese	Mandarin	China			+
78		Lepcha	Lepcha	Bhutan, India, Nepal	+		
79		Qiangic	Qiang	China			+
80	Hmong-Mien	Hmong-Mien	Hmong Njua	China			+
81	Tai-Kadai	Kam-Tai	Mulao	China			+

Macro- Area	Family	Genus	Language	Country	{N, NUM}	{N, NUM, SG}	{N, NUM, CLF}	{N,SG}	Other
82			Thai	Thailand			+		@ {N,CLF}, {NUM,CLF}
83	Austro- Asiatic	Aslian	Semelai	Malaysia			+		
84		Bahnaric	Chrau	Vietnam			+		
85		Khmer	Khmer	Cambodia			+		
86		Nicobarese (Car)	Nicobarese (Car)	India			+		
87		Palau- Khmuic	Khmu'	Laos	@		+		
88		Viet-Muong	Vietnamese	Vietnam			+		@ {N,CLF}
89	Andaman- ese	Great Andaman- ese	Great Andamanese	India	+				
90	Austrones- ian	Atayalic	Atayal	Taiwan	+				
91		Borneo	Begak (Ida'an)	Malaysia			+		
92		Central Malayo- Polynesian	Buru	Indonesia	+				
93			Kambera	Indonesia			+		
94		Oceanic	Erromangan	Vanuatu	+				
95			Kilivila	Papua New Guinea			+		
96			Maori	New Zealand	+				
97			Nalik	Papua New Guinea	+				
98			Rotuman	Fiji	+				

Table A: Sample languages and their structural types of CNNC_{SG}

Macro -Area	Family	Genus	Language	Country	Type		
					{N, NUM}	{N, NUM, SG}	{N, NUM, CLF}
							Other

99		South Halmahera-West New Guinea	Taba	Indonesia			+	
100		Sulawesi	Tukang Besi	Indonesia			+	
101	Creoles-Pidgins		Bislama	Vanuatu	+			
102			Hawai'ian Creole	US (Hawaii)	+			
AUS&NEW	West Papuan	North Halmaheran	Tidore	Indonesia	+		+	
103								
104	East Bird's Head	East Bird's Head	Meyah	Indonesia			+	
105	East Geelvink Bay	East Geelvink Bay	Bauzi	Indonesia	+			
106	Lakes Plain	Lakes Plain	Iau	Indonesia	+			
107	Kwerba	Kwerba	Kwerba	Indonesia	+			
108	Nimboran	Nimboran	Nimboran	Indonesia	+			
109	Sentani	Sentani	Sentani	Indonesia	+			
110	Border	Border	Imonda	Papua New Guinea	+			@ {N,NUM,NPL}
111	Sepik	Ram	Awtuw	Papua New Guinea	+			
112	Lower Sepik-Ramu	Lower Sepik	Yimas	Papua New Guinea		+		
113	Trans-New Guinea	Awju-Dumut	Kombai	Indonesia	+			
114		Binanderean	Binandere	Papua New Guinea	+			
115		Eastern Highlands	Usarufa	Papua New Guinea	+			
116		Fimisterre-Huon	Nabak	Papua New Guinea	+			

Macro -Area	Family	Genus	Language	Country	(N, NUM)	(N, NUM, SG)	(N, NUM, CLF)	(N,SG)	Other
117		Mek	Yale (Kosarek)	Indonesia	+				
118	Marind	South Bird's Head	Inanwatan	Indonesia	+				
119	Eleman	Eleman	Kaki Ae	Papua New Guinea	+				
120	Yele	Yele	Yeli Dnye	Papua New Guinea	+				
121	Sulka	Sulka	Sulka	Papua New Guinea		+			
122	Kuot	Kuot	Kuot	Papua New Guinea	+				
123	Australian	Bunuban	Bunuba	Australia	+				
124		Gaagudju	Gaagudju	Australia	+				
125		Limlingan	Limlingan	Australia	+				
126		Ndjébbana	Ndjébbana	Australia		+			
127		Pama- Nyungan	Arabana	Australia	+				
128			Ngiyambaa	Australia	+				
129		West Barkly	Wambaya	Australia	+				
130	Creoles and Pidgins		Broken	Australia	+				
131			Ndyuka	French Guiana, Suriname	+				
132			Tok Pisin	Papua New Guinea	+				
NAM 133	Na-Dene	Athapaskan	Chipewyan	Canada	+				
134	Haida	Haida	Haida (Skidegate)	Canada			+		
135	Algic	Algonquian	Cree (Plains)	Canada	+				
136		Yurok	Yurok	US			+		

Table A: Sample languages and their structural types of CNNC_{SG}

Macro -Area	Family	Genus	Language	Country	Type		
					{N, NUM}	{N, NUM, SG}	{N, NUM, CLF}
137	Iroquoian	Northern Iroquoian	Mohawk (Akwesasne)	Canada	@		+
138	Siouan	Siouan	Lakota	US	+		
139	Keresan	Keresan	Acoma	US	+		
140	Uto-Aztecan	Aztecan	Nahuatl (Huasteca)	Mexico	+		
141		Cahita	Yaqui	Mexico	+		
142	Salishan	Central Salishan	Halkomelem	Canada	+		
143	Penutian	Tsimshian	Tsimshian	Canada, US			+
144	Oto-Manguean	Chinantecan	Chinantec (Comaltepec)	Mexico	+		
145		Mixtecan	Mixtec (Alacatlalzala)	Mexico	+		
146		Otomian	Ocuilteco	Mexico	+		
147		Pamean	Pame	Mexico	+	@	@
148	Mayan	Mayan	Chontal Maya	Mexico			+
149			Itzaj	Guatemala			+
150			Mam	Guatemala			
151	Creoles and Pidgins		Jamaican Creole	Jamaica	+		
SAM 152	Chibchan	Aruak	Ika	Colombia	+		
153		Guaymí	Buglere	Panama			+
154		Kuna	Kuna	Colombia, Panama			+
155		Paya	Pech	Honduras	+		
156		Talamanca	Teribe	Coasta Rica, Panama			+
157	Barbacoan	Barbacoan	Awa Pit	Colombia, Ecuador	+		
158	Sáliban	Sáliban	Piaroa	Venezuela			+

Macro -Area	Family	Genus	Language	Country	{N, NUM}	{N, NUM, SG}	{N, NUM, CLF}	{N,SG}	Other
159	Tucanoan	Tucanoan	Tucano	Brazil, Colombia			+		
160	Warao	Warao	Warao	Venezuela	+				
161	Yanomam	Yanomam	Sanuma	Brazil, Venezuela	+				
162	Peba-Yaguan	Peba-Yaguan	Yagua	Peru			+		
163	Urarina	Urarina	Urarina	Peru	+				
164	Aymaran	Aymaran	Jaqaru	Peru	+				
165	Arawakan	Arawakan	Tariana	Brazil			+		
166			Warekena	Brazil, Colombia, Venezuela			+		
167	Tupian	Tupi-Guaraní	Urubú-kaapor	Brazil	+				
168	Macro-Ge	Kariri	Kariri	Brazil	+				
169	Chapacura-Wanham	Chapacura-Wanham	Wari'	Brazil					No CNNC _{SG} proper
170	Mura	Mura	Pirahã	Brazil					No CNNC _{SG} proper
171	Tacanan	Tacanan	Cavineña	Bolivia	+				
172	Cayuvava	Cayuvava	Cayuvava	Bolivia	+				
173	Mosetenan	Mosetenan	Mosetén	Bolivia	+				
174	Araucanian	Araucanian	Mapudungun	Chile	+				
175	Creoles and Pidgins		Berberic Dutch Creole	Guyana	+				

Table A: Sample languages and their structural types of CNNC_{SG}

Macro- Area	Family	Genus	Language	Country	1	2	3	4	5	6	7	8
					{N, NUM}	{N, NUM, NSG}	{N, NUM, CLF}	{N, NUM, CLF, NSG}	{N, NUM, OBL, SG}	{N, NUM, OBL, NSG}	{N, NSG}	OTHER
AFR	Khoisan	Central Khoisan	!Ani	Botswana	+							
1												
2	Niger- Congo	Adamawa- Ubangian	Mundang	Chad	+							
3		Bantoid	Lunda	Democratic Republic of Congo	+							
4		Cross River	Kana	Nigeria		+						
5		Dogon	Dogon	Mali	+							
6		Edoid	Degema	Nigeria	+							+ {N,NUM,NSG, DPM}
7		Gur	Koromfe	Burkina Faso, Mali	+							
8		Ijoid	Ijo (Kolokuma)	Nigeria	+							
9		Kru	Grebo	Liberia	+							
10		Northern Atlantic	Wolof	Gambia, Senegal	+							
11		Western Mande	Kpelle	Liberia	+							
12	Nilo- Saharan	Berta	Berta	Ethiopia, Sudan	+							
13		Fur	Fur	Sudan	+							
14		Moru-Ma'di	Ma'di	Sudan, Uganda	+							
15		Nilotic	Lango	Uganda	+							
16		Shabo	Shabo	Ethiopia	+							
17		Songhay	Koyraboro Senni	Mali, Niger	+							
18	Afro- Asiatic	Berber	Berber (Ayt Seghrouchen Middle Atlas)	Morocco	+				+	+		+ {N,NUM,SG}
19		East Chadic	Lele	Chad	+							

Table B: Sample languages and their structural types of CNNC_{NSG}

Macro Area	Family	Genus	Language	Country	1 (N, NUM)	2 (N, NUM, NSG)	3 (N, NUM, CLF)	4 (N, NUM, CLF, NSG)	5 (N, NUM, OBL, SG)	6 (N, NUM, OBL, NSG)	7 (N, NSG)	8 OTHER
38			Gaelic	UK (Scotland)	+	+						
39		Germanic	German	Austria, Germany, Switzerland	@	+						
40			English	United Kingdom	@	+	@					
41			Norwegian	Norway		+						
42		Greek	Greek (Modern)	Greece		+						
43		Indic	Bengali	India, Bangladesh			+					
44			Nepali	Nepal			+					@ {NUM,CLF}
45		Iranian	Persian	Iran	+	+	+					
46		Romance	Catalan	Spain		+						
47			French	France		+				+		
48		Slavic	Russian	Russia	+				+	+		
49	Uralic	Finnic	Finnish	Finland					+			
50			Estonian	Estonia					+			
51		Ugric	Hungarian	Hungary	+		@					
52		Samoyedic	Selkup	Russia	+							
53	Altaic	Mongolic	Mangghuer	China			+					
54			Mongol (Khamnigan)	China, Mongolia	+							
55		Tungusic	Manchu	China	+	+	+					
56			Udihe	Russia	+	+						
57		Turkic	Turkish	Turkey	+		@					
58	Yukaghir	Yukaghir	Yukaghir (Kolyma)	Russia	+	@						
59			Yukaghir (Tundra)	Russia	+							

Table B: Sample languages and their structural types of CNNC_{NSG}

Macro-Area	Family	Genus	Language	Country	Type									
					1	2	3	4	5	6	7	8		
					{N, NUM}	{N, NUM, NSG}	{N, NUM, CLF}	{N, NUM, CLF, NSG}	{N, NUM, OBL, SG}	{N, NUM, OBL, NSG}	{N, NSG}	OTHER		
60	Yeniseian	Yeniseian	Ket	Russia		+								
61	Chukotko-Kamchatkan	Northern Chukotko-Kamchatkan	Alutor	Russia		+								
62	Nivkh	Nivkh	Nivkh	Russia			+							
63	Ainu	Ainu	Ainu	Japan	+		@							
64	Japanese	Japanese	Japanese	Japan			+							
65	Korean	Korean	Korean	Korea			+	@						
66	Northwest Caucasian	Northwest Caucasian	Abkhaz	Georgia	+	+								
67	Nakh-Daghestanian	Avar-Andic-Tsezic	Hunzib	Russia	+	@								
68		Lezgian	Lezgian	Russia	+									
69	Kartvelian	Kartvelian	Georgian	Georgia	+									
70	Burushaski	Burushaski	Burushaski	Pakistan		+								
71	Nahali	Nahali	Nahali	India		+								
72	Dravidian	Central Dravidian	Kolami	India		+	+							
73		Northern Dravidian	Malto	India			+	+						
74		Southern Dravidian	Tamil	India	+	+								
75	Austro-Asiatic (in Eurasia)	Munda	Bhumij	India	+									
76			Korku	India	+	+							+	
77			Mundari	India		+								
78			Santali	India	+		+						+	
SEA&OCE	Sino-Tibetan	Bai	Bai	China			+							
79														
80		Bodic	Camling	Nepal			+	+						
81			Chantyal	Nepal			+	@						

Macro-Area	Family	Genus	Language	Country	1	2	3	4	5	6	7	8
					(N, NUM)	(N, NUM, NSG)	(N, NUM, CLF)	(N, NUM, CLF, NSG)	(N, NUM, OBL, SG)	(N, NUM, OBL, NSG)	(N, NSG)	OTHER
82			Tamang	Nepal	+							
83			Tibetan (Standard Spoken)	China (Tibet)	+							
84		Burmese-Lolo	Akha	China, Myanmar, Thailand		+						
85		Chinese	Mandarin	China		+						
86		Kuki-Chin-Naga	Meithei	India	+							
87		Lepcha	Lepcha	Bhutan, India, Nepal	+							
88		Nungish	Dulong	China			+					
89		Qiangic	Qiang	China			+					
90	Hmong-Mien	Hmong-Mien	Hmong Njua	China			+					
91	Tai-Kadai	Kam-Tai	Mulao	China			+					
92			Thai	Thailand			+					@ {NUM,CLF}
93	Austro-Asiatic	Aslian	Semelai	Malaysia			+					
94		Bahnaric	Sapuan	Laos			+					
95		Khmer	Khmer	Cambodia	+		+					
96		Palaung-Khmuic	Khmu'	Laos			+					
97		Viet-Muong	Vietnamese	Vietnam			+					@ {NUM,CLF}
98	Andamanese	Great Andamanese	Great Andamanese	India	+							
99	Austronesian	Atayalic	Atayal	Taiwan	+							
100		Borneo	Begak (I'da'an)	Malaysia	@		+					

Table B: Sample languages and their structural types of CNNC_{NSG}

Macro-Area	Family	Genus	Language	Country	Type								
					1	2	3	4	5	6	7	8	
					{N, NUM}	{N, NUM, NSG}	{N, NUM, CLF}	{N, NUM, CLF, NSG}	{N, NUM, OBL, SG}	{N, NUM, OBL, NSG}	{N, NSG}	OTHER	
101		Central Malayo-Polynesian	Buru	Indonesia	+	+	+						
102			Kambera	Indonesia			+						+ {N, NUM, RMS}
103			Leti	Indonesia	+	+							
104		Oceanic	Erromangan	Vanuatu	+	@							
105			Futuna-Aniwa	Vanuatu		+					+		
106			Kilivila	Papua New Guinea			+	+					
107			Maori	New Zealand									+ {N, NUM, NUMPCL}, @ {N, NUM, PNUM}
108			Nalik	Papua New Guinea	+	+					+		
109			Rotuman	Fiji	+								
110			Tuvaluan	Indonesia	+	@							@ {N, NUM, NUMPCL, NSG}
111		South Halmahera-West New Guinea	Taba	Indonesia			+	+					
112		Sundic	Madurese	Indonesia	+								
113		Sulawesi	Tukang Besi	Indonesia			+						
114	Creoles-Pidgins		Bislama	Vanuatu	+		@						
115			Hawaiian Creole	US (Hawaii)			+						
AUS&NEW 116	West Papuan	Hatam	Hatam	Indonesia	+		+						

Area	Country	Language	Country	Structural types										
				1 (N, NUM)	2 (N, NUM, NSG)	3 (N, NUM, CLF)	4 (N, NUM, CLF, NSG)	5 (N, NUM, OBL, SG)	6 (N, NUM, OBL, NSG)	7 (N, NSG)	8 OTHER			
117		Kebar	Indonesia			+								
118		North Halmaheran	Indonesia	+		+								
119		North Central Bird's Head	Indonesia			+								
120	East Bird's Head	East Bird's Head	Indonesia			+								
121	Lower Mambera- mo	Lower Mambera- mo	Indonesia	+								+		
122	East Geelvink Bay	East Geelvink Bay	Indonesia		+									
123	Lakes Plain	Lakes Plain	Indonesia	+										
124	Kwerba	Kwerba	Indonesia	+										
125	Nimboran	Nimboran	Indonesia	+										
126	Sentani	Sentani	Indonesia	+										
127	Sko	Western Sko	Papua New Guinea	+										
128	Border	Border	Papua New Guinea	+								+	@ (N,NUM,NPL)	
129	Torricelli	Wapei-Palei	Papua New Guinea		+									
130		Kombio- Arapesh	Papua New Guinea		+									
131	Sepik	Ram	Papua New Guinea	+	+									
132		Middle Sepik	Papua New Guinea	+										
133	Lower Sepik-Ramu	Lower Sepik	Papua New Guinea		+									
134	Trans-New Guinea	Awju- Dumut	Indonesia	+										

Table B: Sample languages and their structural types of CNN_{NSG}

Area	Family	Genus	Language	Country	1 (N, NUM)	2 (N, NUM, NSG)	3 (N, NUM, CLF)	4 (N, NUM, CLF, NSG)	5 (N, NUM, OBL, SG)	6 (N, NUM, OBL, NSG)	7 (N, NSG)	8 OTHER
157	Creoles and Pidgins		Broken	Australia	+							
158			Tok Pisin	Papua New Guinea	+							
NAM												
159	Eskimo-Aleut	Eskimo-Aleut	Yup'ik	US		+					+	
160	Na-Dene	Athapaskan	Carrier	Canada			+					
161			Chipewyan	Canada								
162			Sarcee	Canada			+					
163		Tlingit	Tlingit	US								
164	Haida	Haida	Haida	Canada			@					
165	Algic	Algonquian	Cree (Plains)	Canada			+					
166		Yurok	Yurok	US				+				
167	Iroquoian	Northern Iroquoian	Mohawk (Akwasasne)	Canada			+				+	
168	Yuchi	Yuchi	Yuchi	US			+					
169	Muskogean	Muskogean	Koasati	US			+					
170	Chitimacha	Chitimacha	Chitimacha	US								
171	Caddoan	Caddoan	Wichita	US								
172	Siouan	Siouan	Lakhota	US								
173	Kiowa-Tanoan	Kiowa-Tanoan	Jemez	US								+ {N,NUM,INV}
174	Keresan	Keresan	Acoma	US								
175	Zuni	Zuni	Zuni	US								+ {N,NUM,SG}
176	Uto-Aztecan	Aztecan	Nahuatl (Huasteca)	Mexico			+					
177		Cahita	Yaqui	Mexico								
178	Wakashan	Southern Wakashan	Nuuchahnulth	Canada			+					
179	Salishan	Central Salishan	Halkomelem (Musqueam)	Canada				+				
180	Chimakuan	Chimakuan	Quileute	US								
181	Kutenai	Kutenai	Kutenai	US								

Table B: Sample languages and their structural types of CNNC_{NSG}

Area	Language	Country	1 (N, NUM)	2 (N, NUM, NSG)	3 (N, NUM, CLF)	4 (N, NUM, CLF, NSG)	5 (N, NUM, OBL, SG)	6 (N, NUM, OBL, NSG)	7 (N, NSG)	8 OTHER
204	Sálitban	Sálitban			+	@				
205	Tucanoan	Tucanoan			+					
206	Warao	Warao	+			+				
207	Yanomam	Yanomam		+						
208	Waorani	Waorani							+	
209	Zaparoan	Zaparoan	+							
210	Peba-Yaguan	Peba-Yaguan			+				+	
211	Urarina	Urarina			+					
212	Chimúan	Chimúan	+							
213	Panoan	Panoan	+							
214	Quechuan	Quechuan	+		+					
215	Aymaran	Aymaran	+							
216	Arawakan	Arawakan				+				
217		Warekena	+							
218	Tupian	Ramarama	+							
219		Tupi-Guaraní	+							
220	Trumai	Trumai	+							
221	Kwaza	Kwaza			+					
222	Chapacura-Wanhan	Chapacura-Wanhan								No CNNC _{NSG} proper
223	Mura	Mura								No CNNC _{NSG} proper
224	Arauan	Arauan	+							
225	Tacanan	Tacanan	+							
226	Cayuvava	Cayuvava	+							
227	Mosetenan	Mosetenan	+							

Table B: Sample languages and their structural types of CNNC_{NSG}

Macro-Area	Family	Genus	Language	Country	Type							
					1	2	3	4	5	6	7	8
					{N, NUM}	{N, NUM, NSG}	{N, NUM, CLF}	{N, NUM, CLF, NSG}	{N, NUM, OBL, SG}	{N, NUM, OBL, NSG}	{N, NSG}	OTHER
228	Guaicuruan	Guaicuruan	Mocoví	Argentina	+	+						
229			Pilagá	Argentina		+						
230	Araucanian	Araucanian	Mapudun- gun	Chile	+							
231	Woorani	Woorani	Woorani	Ecuador							+	
232	Creoles and Pidgins		Berbice Dutch Creole	Guyana	+	@						
233			Ndyuka	French Guiana, Suriname	+							

Appendix 2 Lists of Languages

Appendix 2 provides lists of languages. Appendix 2.1 provides a list of sample languages used in the thesis. The sample languages are listed alphabetically by the name. In Appendix 2.2 – Appendix 2.5, lists of languages corresponding to the figures in the summary tables in Chapter 6 and Chapter 7 are offered. Appendix 2.2 and Appendix 2.3 show lists of languages corresponding to the figures in Table 6.1 and Table 6.2 respectively. Appendix 2.4 provides a list of languages corresponding to the figures in Table 7.1. Finally, Appendix 2.5 provides lists of languages corresponding to the figures in Table 7.9-Table 7.13 respectively.

Appendix 2.1 List of sample languages

Abkhaz	Awtuw	Buru
Abun	Bai	Burushaski
Acoma	Barasano	Camling
Ainu	Basque	Carrier
Akha	Bauzi	Catalan
Albanian	Begak-Ida'an	Cavineña
Alutor	Bengali	Cayuvava
Ambulas	Berber (Ayt	Chaha
Amele	Seghrouchen Middle	Chantyal
Angolar	Atlas)	Chimila
//Ani	Berbice Dutch Creole	Chinantec
Arabana	Berta	(Comaltepec)
Arabic (Modern	Bhumij	Chipewyan
Standard)	Binandere	Chitimacha
Arapesh	Bislama	Chontal Maya
Armenian (Eastern)	Bribri	Chrau
Armenian (Western)	Broken	Cree (Plains)
Atayal	Buglere	Degema
Awa Pit	Bunuba	Diyari

Dogon	Itzaj	Lele
Dulong	Jamaican Creole	Lepcha
Dumo	Japanese	Leti
Ejagham	Jaqaru	Lezgian
English	Jarawara	Limilngan
Erromangan	Jemez	Lithuanian
Estonian	Kaki Ae	Lunda
Finnish	Kambera	Maale
French	Kana	Ma'di
Fur	Kanuri	Madurese
Futuna-Aniwa	Kariri	Maltese
Gaagudju	Karó (Arára)	Malto
Gaelic (Scots)	Kayardild	Mam
Georgian	Ket	Manchu
German	Khmer	Mandarin
Great Andamanese	Khmu'	Mangghuer
Grebo	Kilivila	Maori
Greek (Modern)	Koasati	Mapudungun
Haida	Kolami	Matsés
Halkomelem	Kombai	Maung
Hatam	Korean	Meithei
Hausa	Korku	Meyah
Hawaiian Creole	Koromfe	Miskito
Hebrew (Modern)	Koyraboro Senni	Mixtec (Alacatlazala)
Hmong Njua	Kpelle	Mochica
Hungarian	Kuna	Mocoví
Hunzib	Kuot	Mohawk
Iau	Kutenai	Mongol (Khamnigan)
Ijo (Kolokuma)	Kwazá	Mosetén
Ika	Kwerba	Mpur
Imonda	Lakhota	Mulao
Inanwatan	Lango	Mundang
Iquito	Latvian	Mundari
Irish	Lavukaleve	Nabak

Nahali	Selkup	Warao
Nahuatl (Huasteca)	Semelai	Warekena
Nalik	Sentani	Warembori
Ndjébbana	Shabo	Wari'
Ndyuka	Somali	Welsh
Nepali	Sulka	Wichita
Nez Perce	Taba	Wolof
Ngiyambaa	Tamang	Yagua
Nicobarese (Car)	Tamil	Yale (Kosarek)
Nigerian Pidgin	Tariana	Yaqui
Nimboran	Teribe	Yeli Dnye
Nivkh	Thai	Yidiny
Norwegian	Tibetan (Standard	Yimas
Nubi	Spoken)	Yuchi
Nuuchahnulth	Tidore	Yukaghir (Kolyma)
Ocuilteco	Tiwi	Yukaghir (Tundra)
Olo	Tlingit	Yup'ik (Central)
Olutec	Tok Pisin	Yurok
Orig	Tol	Zuni
Pame	Totonac (Misantla)	
Pech	Trumai	
Persian	Tsimshian (Coast)	
Piaroa	Tucano	
Pilagá	Tukang Besi	
Pirahã	Turkish	
Qiang	Tuvaluan	
Quechua (Huallaga)	Udihe	
Quileute	Urarina	
Rotuman	Urubú-Kaapor	
Russian	Usarufa	
Santali	Vietnamese	
Sanuma	Wambaya	
Sapuan	Waorani	
Sarcee	Wappo	

Appendix 2.2 Lists of languages for Table 6.1 (CNNC_{SG})

1. {N,NUM} [117 languages]

Abkhaz	English	Korku
Acoma	Erromangan	Koyraboro Senni
Ainu	Estonian	Kpelle
Albanian	Finnish	Kuot
Angolar	French	Kwerba
Arabana	Fur	Lakhota
Arabic (Modern Standard)	Gaagudju	Lango
Atayal	Gaelic (Scots)	Lele
Awa Pit	Georgian	Lepcha
Awtuw	German	Lezgian
Basque	Great Andamanese	Limilngan
Bauzi	Grebo	Lithuanian
Berbice Dutch Creole	Greek (Modern)	Maale
Berta	Hausa	Ma'di
Bhumij	Hawaiian Creole	Maltese
Binandere	Hebrew (Modern)	Mam
Bislama	Hungarian	Manchu
Broken	Iau	Maori
Bunuba	Ijo (Kolokuma)	Mapudungun
Buru	Ika	Mixtec (Alacatlatzala)
Burushaski	Imonda	Mongol (Khamnigan)
Catalan	Inanwatan	Mosetén
Cavineña	Irish	Mundari
Cayuvava	Jamaican Creole	Nabak
Chaha	Jaqaru	Nahali
Chinantec(Comalte- pec)	Kaki Ae	Nahuatl (Huasteca)
Chipewyan	Kanuri	Nalik
Cree (Plains)	Ket	Ndyuka
	Kolami	Ngiyambaa
	Kombai	Nigerian Pidgin

Nimboran	Selkup	Wambaya
Norwegian	Sentani	Warao
Nubi	Somali	Welsh
Ocuilteco	Tamang	Wolof
Pame	Tok Pisin	Yale (Kosarek)
Pech	Turkish	Yaqui
Persian	Udihe	Yelî Dnye
Rotuman	Urarina	Yukaghir (Kolyma)
Russian	Urubú-Kaapor	Yukaghir (Tundra)
Sanuma	Usarufa	

2. *{N,NUM,SG}* [10 languages]

Alutor	Koromfe	Sulka
//Ani	Lunda	Yimas
Degema	Ndjébbana	
Ejagham	Orig	

3. *{N,NUM,CLF}* [39 languages]

Akha	Kana	Qiang
Begak-Ida'an	Khmu'	Semelai
Bengali	Kilivila	Taba
Buglere	Korean	Tariana
Camling	Kuna	Teribe
Chantyal	Malto	Thai
Chontal Maya	Mandarin	Tsimshian (Coast)
Chrau	Mangghuer	Tucano
Haida	Meyah	Tukang Besi
Hmong Njua	Mulao	Vietnamese
Itzaj	Nepali	Warekena
Japanese	Nivkh	Yagua
Kambera	Piaroa	Yurok

4. *{N,NUM} + {N,NUM,CLF}* [5 languages]

Halkomelem

Khmer

Tidore

Kariri

Nicobarese (Car)

5. *Other* [2 languages]

{N,SG} [1 language]

Mohawk

{N,NUM,SG} + {N,NUM,OBL,SG} [1 language]

Berber (Ayt

Seghrouchen Middle

Atlas)

6. *None* [2 languages]

Pirahã

Wari'

Appendix 2.3 Lists of languages for Table 6.2 (CNNC_{NSG})

1. {N,NUM} [78 languages]

Acoma	Iau	Ngiyambaa
Ainu	Ijo (Kolokuma)	Nimboran
Ambulas	Ika	Nubi
Amele	Iquito	Ocuilteco
Angolar	Jaqaru	Pech
Atayal	Karó (Arára)	Quileute
Awa Pit	Kombai	Rotuman
Basque	Koyraboro Senni	Selkup
Berbice Dutch Creole	Kpelle	Sentani
Bhumij	Kutenai	Shabo
Bislama	Kwerba	Tamang
Broken	Lakhota	Tibetan (Standard Spoken)
Cavineña	Lepcha	Tlingit
Cayuvava	Lezgian	Tok Pisin
Chaha	Ma'di	Turkish
Chimila	Madurese	Tuvaluan
Chinantec (Comaltepec)	Mam	Urubú-Kaapor
Chipewyan	Mapudungun	Wappo
Chitimacha	Matsés	Warao
Dogon	Maung	Wichita
Dumo	Meithei	Yidiny
Erromangan	Miskito	Yukaghir (Kolyma)
Georgian	Mixtec (Alacatlalzala)	Yukaghir (Tundra)
Great Andamanese	Mochica	
Haida	Mongol (Khamnigan)	
Hungarian	Nabak	
Hunzib	Ndjébbana	
	Ndyuka	

2. {N,NUM,NSG} [36 languages]

Albanian	German	Nahali
Alutor	Grebo	Norwegian
//Ani	Greek (Modern)	Olo
Arapesh	Hawaiian Creole	Pilagá
Bauzi	Inanwatan	Sanuma
Berta	Kaki Ae	Sarcee
Bribri	Ket	Sulka
Burushaski	Koasati	Wolof
Catalan	Koromfe	Yaqui
Cree (Plains)	Latvian	Yimas
English	Lunda	
Fur	Mundang	
Gaagudju	Mundari	

3. {N,NUM,CLF} [33 languages]

Abun	Korean	Qiang
Akha	Kuna	Sapuan
Bai	Kwazá	Semelai
Begak-Ida'an	Mandarin	Teribe
Bengali	Mangghuer	Thai
Chantyal	Meyah	Totonac (Misantla)
Chontal Maya	Mpur	Tukang Besi
Dulong	Mulao	Vietnamese
Hmong Njua	Nepali	Yurok
Japanese	Nicobarese (Car)	
Kana	Nivkh	
Khmu'	Piaroa	

4. Other [10 languages]

{N,NUM,OBL,SG} [2 languages]

Estonian Finnish

{N,NSG} [3 languages]

Bunuba Diyari Waorani

{N,NUM,NUMPCL} [1 language]

Maori

{N,NUM,CLF,NSG} [2 languages]

Tariana Tsimshian (Coast)

{N,NUM,INV} [1 language]

Jemez

{N,NUM,SG} [1 language]

Zuni

5. {N,NUM} + {N,NUM,NSG} [33 languages]

Abkhaz	Lavukaleve	Olutec
Armenian (Western)	Lele	Pame
Awtuw	Leti	Quechua (Huallaga)
Binandere	Limilngan	Tamil
Gaelic (Scots)	Maale	Tiwi
Hausa	Maltese	Tol
Hebrew (Modern)	Mocoví	Trumai
Irish	Mosetén	Udihe
Jamaican Creole	Nahuatl (Huasteca)	Warekena
Jarawara	Nigerian Pidgin	Yale (Kosarek)
Lango	Nuuchahnulth	Yuchi

6. $\{N, NUM\} + \{N, NUM, CLF\}$ [3 languages]

Hatam	Khmer	Tidore
-------	-------	--------

7. $\{N, NUM, CLF\} + \{N, NUM, CLF, NSG\}$ [8 languages]

Barasano	Itzaj	Taba
Camling	Kilivila	Urarina
Carrier	Malto	

8. $\{N, NUM, (NSG)\} + \{N, NUM, OBL, SG/NSG\}$ [5 languages]

Lithuanian	Somali	French
Russian	Welsh	

9. $\{N, NUM, (NSG)\} + \{N, NSG\}$ [12 languages]

Arabana	Korku	Wambaya
Futuna-Aniwa	Kuot	Warembori
Imonda	Mohawk	Yelí Dnye
Kayardild	Nalik	Yup'ik (Central)

10. *Mixed* [13 languages]

Arabic (Modern Standard)	Buru	Nez Perce
Armenian (Eastern)	Degema	Persian
Berber (Ayt Seghrouchen Middle Atlas)	Halkomelem	Santali
	Kambera	Yagua
	Kolami	
	Manchu	

11. *None* [2 languages]

Pirahã	Wari'	
--------	-------	--

Appendix 2.4

Lists of languages for Table 7.1

+Plural/+Gender	Latvian	Armenian (Eastern)
[54 languages]	Lavukaleve	Bagirmi
Abkhaz	Lele	Basque
Alamblak	Luvale	Bawm
Apurinã	Macushi	Brahui
Arabic (Egyptian)	Maltese	Chamorro
Babungo	Mixtec (Chalcatongo)	Chukchi
Barasano	Miya	Chuvash
Bininj Gun Wok	Mosetén	Comanche
Burushaski	Ojibwa (Eastern)	Coos (Hanis)
Chichewa	Oromo (Harar)	Epena Pedee
Cree (Plains)	Panjabi	Erromangan
Dyirbal	Qafar	Evenki
English	Retuarã	Finnish
French	Russian	Georgian
German	Spanish	Greenlandic (West)
Grebo	Supyire	Guaraní
Greek (Modern)	Swahili	Hatam
Hausa	Tagalog	Hungarian
Hebrew (Modern)	Tamil	Indonesian
Hindi	Tiwi	Kanuri
Hixkaryana	Wambaya	Khalkha
Hunzib	Wardaman	Khmer
Icelandic	Yimas	Koasati
Iraqw	Zulu	Kobon
Jul'hoan		Kutenai
Kannada	+Plural/-Gender	Lango
Ket	[55 languages]	Lepcha
Kolami	Acoma	Lezgian
Koromfe	Ainu	Mandarin
Lak	Amele	Maori

Maricopa	Tibetan (Standard	Imonda
Mokilese	Spoken)	Kayardild
Ngiyambaa	Turkish	Mapudungun
Nivkh	Tuvaluan	Usan
Nubian (Dongolese)	Vietnamese	Yidiny
Palauan	Yaqui	-Plural/+Gender
Pomo (Central)	Yoruba	[3 languages]
Quechua (Imbabura)	Yup'ik (Central)	Maybrat
Rama	Zuni	Pirahã
Rapanui		Tidore
Shipibo	-Plural/-Gender	
Konibo	[7 languages]	
Suena	Chimariko	
Taba	Hmong Njua	

Appendix 2.5 Lists of languages for Table 7.9-Table 7.13

Appendix 2.5.1 $\{N, NUM, NSG, (X)\}$ and $[+NC]$

Languages	Genera
AFR [11 genera]	
1. //Ani	Central Khoisan
2. Arabic (MS)	Semitic
3. Berber (ASMA)	Berber
4. Degema	Edoi
5. Grebo	Kru
6. Hausa	West Chadic
7. Hebrew	Semitic* ⁵²
8. Koromfe	Gur
9. Lele	East Chadic
10. Lunda	Bantoid
11. Maltese	Semitic*
12. Somali	Eastern Cushitic
13. Wolof	Northern Atlantic
EUR [16 genera]	
1. Abkhaz	Northwest Caucasian
2. Albanian	Albanian
3. Burushaski	Burushaski
4. Catalan	Romance
5. French	Romance*
6. Gaelic (Scots)	Celtic
7. German	Germanic
8. Greek	Greek
9. Hunzib	Avar-Andic-Tsezic
10. Irish	Celtic*

⁵² The genera with an asterisk are the genera which are repeated. Since one genus is counted only once, the repeated genera will not be counted.

11. Ket	Yeniseian
12. Kolami	Central Dravidian
13. Latvian	Baltic
14. Lithuanian	Baltic*
15. Malto	Northern Dravidian
16. Mundari	Munda
17. Nahali	Nahali
18. Norwegian	Germanic*
19. Russian	Slavic
20. Tamil	Southern Dravidian
21. Welsh	Celtic*

AUS-NEW [9 genera]

1. Arapesh	Kombio-Arapesh
2. Gaagudju	Gaagudju
3. Inanwatan	South Bird's Head
4. Kuot	Kuot
5. Lavukaleve	Solomons East Papuan
6. Olo	Wapei-Palei
7. Tiwi	Tiwian
8. Wambaya	West Barkly
9. Yimas	Lower-Sepik

NAM [3 genera]

1. Cree (Plains)	Algonquian
2. Mohawk	Northern Iroquoian
3. Yuchi	Yuchi

SAM [6 genera]

1. Barasano	Tucanoan
2. Jarawara	Arauan
3. Mosetén	Mosetenan

- | | |
|-------------|-------------|
| 4. Pilagá | Guaicuruan |
| 5. Tariana | Arawakan |
| 6. Warekena | Arawakan* |
| 7. Yagua | Peba-Yaguan |

Appendix 2.5.2 *+{N,NUM,NSG,(X)} and [-NC]*

AFR [5 genera]

- | | |
|--------------------|---------------------|
| 1. Berta | Berta |
| 2. Fur | Fur |
| 3. Lango | Nilotic |
| 4. Mundang | Adamawa-Ubangian |
| 5. Nigerian Pidgin | Creoles and Pidgins |

EUR [6 genera]

- | | |
|-----------------------|-----------|
| 1. Armenian (Eastern) | Armenian |
| 2. Armenian (Western) | Armenian* |
| 3. English | Germanic |
| 4. Korean | Korean |
| 5. Manchu | Tungusic |
| 6. Persian | Iranian |
| 7. Udihe | Tungusic* |
| 8. Yukaghir (Kolyma) | Yukaghir |

SEA & OCE [3 genera]

- | | |
|--------------------|---------------------------------|
| 1. Erromangan | Oceanic |
| 2. Hawaiian Creole | Creoles and Pidgins |
| 3. Kilivila | Oceanic* |
| 4. Nalik | Oceanic* |
| 5. Taba | South Halmahera-West New Guinea |
| 6. Tuvaluan | Oceanic* |

AUS-NEW [3 genera]

- | | |
|----------|---------------------------------|
| 1. Awtuw | Ram |
| 2. Sulka | Sulka |
| 3. Taba | South Halmahera-West New Guinea |

NAM [10 genera]

- | | |
|--------------------------|---------------------|
| 1. Bribri | Talamanca |
| 2. Jamaican Creole | Creoles and Pidgins |
| 3. Koasati | Muskogean |
| 4. Nahuatl (Hausteca) | Aztecán |
| 5. Nez Perce | Sahaptian |
| 6. Nuuchahnulth (Nootka) | Southern Wakashan |
| 7. Tol | Tol |
| 8. Tsimshian (Coast) | Tsimshian |
| 9. Yaqui | Cahita |
| 10. Yup'ik | Eskimo-Aleut |

SAM [4 genera]

- | | |
|-------------------------|---------------------|
| 1. Berbice Dutch Creole | Creoles and Pidgins |
| 2. Quechua (Huallaga) | Quechuan |
| 3. Trumai | Trumai |
| 4. Urarina | Urarina |

Appendix 2.5.3 *-{N,NUM,NSG,(X)} and [+NC]*

AFR [1 genera]

- | | |
|----------|-------|
| 1. Dogon | Dogon |
|----------|-------|

EUR [1 genera]

- | | |
|-----------|-------|
| 1. Bhumij | Munda |
|-----------|-------|

SEA & OCE [1 genera]

- | | |
|---------------------|------------|
| 1. Nicobarese (Car) | Nicobarese |
|---------------------|------------|

AUS-NEW [4 genera]

- | | |
|--------------|--------------|
| 1. Diyari | Pama-Nyungan |
| 2. Maung | Iwaidjan |
| 3. Ndjébbana | Ndjébbana |
| 4. Tidore | Tidore |

NAM [1 genera]

- | | |
|-------------|-----------|
| 1. Quileute | Chimakuan |
|-------------|-----------|

*Appendix 2.5.4 -{N,NUM,NSG,(X)} and [-NC]***AFR [5 genera]**

- | | |
|--------------------|----------------------|
| 1. Angolar | Creoles and Pidgins |
| 2. Ijo (Kolokuma) | Ijoid |
| 3. Kana | Cross-River |
| 4. Koyraboro Senni | Songhay |
| 5. Kpelle | Western Mande |
| 6. Nubi | Creoles and Pidgins* |

EUR [13 genera]

- | | |
|--------------|------------|
| 1. Ainu | Ainu |
| 2. Basque | Basque |
| 3. Bengali | Indic |
| 4. Estonian | Finnic |
| 5. Finnish | Finnic* |
| 6. Georgian | Kartvelian |
| 7. Hungarian | Ugric |
| 8. Japanese | Japanese |

9. Lezgian	Lezgiic
10. Mongol (Khamnigan)	Mongolic
11. Nivkh	Nivkh
12. Selkup	Samoyedic
13. Turkish	Turkic
14. Yukaghir (Tundra)	Yukaghir

SEA & OCE [20 genera]

1. Akha	Burmese-Lolo
2. Atayal	Atayalic
3. Bai	Bai
4. Begak (Ida'an)	Borneo
5. Bislama	Creoles and Pidgins
6. Great Andamese	Great Andamanese
7. Hmong Njua	Hmong-Mien
8. Kambera	Central Malayo-Polynesian
9. Khmer	Khmer
10. Lepcha	Lepcha
11. Madurese	Sundic
12. Mandarin	Chinese
13. Maori	Oceanic
14. Meithei	Kuki-Chin-Naga
15. Mulao	Kam-Tai
16. Qiang	Qiangic
17. Rotuman	Oceanic*
18. Semelai	Aslian
19. Thai	Kam-Tai*
20. Tibetan (Standard Spoken)	Bodic
21. Tukang Besi	Sulawesi
22. Vietnamese	Viet-Muong

AUS-NEW [9 genera]

1. Amele	Madang
2. Arabana	Pama-Nyungan
3. Broken	Creoles and Pidgins
4. Bunuba	Bunuban
5. Hatam	Hatam
6. Imonda	Border
7. Kayardild	Tangkic
8. Kombai	Awju-Dumut
9. Ngiyambaa	Pama-Nyungan*
10. Sentani	Sentani
11. Tok Pisin	Creoles and Pidgins*
12. Yidiny	Pama-Nyungan*

NAM [9 genera]

1. Acoma	Keresan
2. Chitimacha	Chitimacha
3. Haida	Haida
4. Kutenai	Kutenai
5. Lakhota	Siouan
6. Wappo	Wappo
7. Wichita	Caddoan
8. Yurok	Yurok
9. Zuni	Zuni

SAM [14 genera]

1. Awa Pit	Barbacoan
2. Cayuvava	Cayuvava
3. Chimila	Chimila
4. Chontal Maya	Mayan
5. Ika	Aruak
6. Iquito	Zaparoan

7. Jaqaru	Aymaran
8. Karó	Ramarama
9. Kwazá	Kwaza
10. Mapudungun	Araucanian
11. Miskito	Misumalpan
12. Ndyuka	Creoles and Pidgins
13. Urubú-Kaapor	Tupi-Guaraní
14. Warao	Warao

Appendix 3 Survey of Reference Grammars and Data⁵³

Appendix 3 provides the summarized grammatical information of sample languages with respect to structural patterns of CNNCs, number systems and classifier systems. The information is mainly extracted from the survey of reference grammars and summarized by using the format as shown below. In this appendix, the sample languages are listed alphabetically by name.

Language:

Family/Genus:

Country/ Macro Area:

Reference(s):

CNNCs

1. Structural patterns of $CNNC_{SG}$:
2. Restrictions on the patterns in (1) above:
3. Structural patterns of $CNNC_{NSG}$:
4. Restrictions on the patterns in (3) above:

Number systems

5. Number distinction:
6. Obligatoriness of number marking:
7. Non-numeral quantifiers:

Classifier systems

8. Noun classes:
9. Numeral classifiers:

⁵³ Note that 1/5 of the current sample of $CNNC_{NSG}$ was also used for the author's MSc by Research dissertation submitted to the department of Theoretical and Applied Linguistics, University of Edinburgh, in September 2004. It is emphasized hereby that the vast majority of the analysis in this PhD thesis is entirely new.

There are various types of answers to the questions in the survey of reference grammars. Below is a list of answers to some questions and what may be inferred from them.

1. Answers to Question (2) and Question (4): Restrictions on the patterns

- *No restrictions*

This answer means that it is stated clearly in the source that the structural patterns in Question (1) and Question (3) are used without restrictions. In this case, the reference of the source is given in parentheses.

- *No known restrictions*

This answer means that no explicit information about restrictions of the structural patterns in question is found in the source. In many cases, the example of structural pattern is presented in the source without any information about the restrictions of use. It is possible that there is a restriction of use in practice. However, the issue of restrictions is not focused on in the source.

- *Used generally with no known restrictions*

This answer is used in the case in which the structural patterns can be observed in several places in the source and in which the pattern is used generally with various classes of nouns. Still, the information about restrictions of these patterns is not explicitly stated. However, the references where the examples of general use can be observed will be given in parentheses.

- *{A,B,C}: Used with X only*

This answer means that the structural pattern of {A,B,C} is used with X context, for example, when the pattern is used in literary works only. The patterns of CNNCs which are regarded as subsidiary are remarked in the survey as “*hence @*”. (@ is the symbol representing a subsidiary mode). Below are examples of cases regarded as subsidiary or @:

(1) It is stated clearly in the source that {A,B,C} is occasional or rare or less frequent than the other pattern(s).

(2) {A,B,C} is used in colloquial forms only, whereas the other structural pattern can be used in both colloquial and written forms.

(3) {A,B,C} is used in literary works only, whereas the other structural patterns can be used in wider genres

(4) {A,B,C} is used with a small class of nouns, e.g. nouns denoting domestic animals.

(5) {A,B,C} is used only when accompanied by noun modifiers.

In some cases, the restrictions involve numerals, that is, a particular pattern is used with a particular number. For example, French {N,NUM,OBL,NSG} is used only with the numerals *million* 'million' and *milliard* 'billion'. Since in many languages such as Gaelic (Scots) or Russian the use of numerals is quite complicated, it is quite difficult to say that the particular number is less used. Therefore, any numerals are treated though as they are used equally.

2. Answers to Question (5): Number distinction

- *SG/DU/TRI/PAU/PL*

This answer means that the language has the number distinction of singular, dual, trial, paucal and plural. The source of reference from which the data obtained will be given in parentheses.

- *No distinction*

This answer means that the language does not have number distinction (on nouns) at all. The source of reference from which the data obtained will be given in parentheses. If the reference is indicated as *passim*, it means that the information is inferred or elicited by the thesis author. That is, the thesis author did not find evidence showing that the language uses number distinction by any means (cf. §5.1).

- *No information*

The information about number distinction was not available to the thesis author. The problem occurs mostly when the source is not a reference grammar, but an article or a book chapter.

3. Answers to Question (6): Obligatoriness of number marking

- *Obligatory*

This answer means that the system of number marking on nouns is generally obligatory at least for most nouns.

- *Optional*

This answer means that the system of number marking is optional.

- *See (5)*

This answer means that the language does not have a number-marking system on nouns and so the question about obligatoriness of number marking is inapplicable. In such a case, the reader is referred to the answer to Question (5).

- *No information*

This answer means that the information on obligatoriness of number marking was not available to the thesis author.

Note that in the case that the source of reference is indicated as (*WALS*), this means that the answer to the question was taken from the database made by Haspelmath (2005). The reader is referred to the *Interactive Reference Tool of the World Atlas of Language Structures* on the feature of *occurrence of nominal plurality* for further information such as the actual source of reference.

4. Answers to Question (7): Non-numeral quantifiers

The answer to Question (7) is to show various kinds of non-numeral quantifiers, typically lexical quantifiers (LQ) (e.g. English *many*), illustrated by examples with glosses. Non-numeral quantifiers can be realised by reduplication, third person plural pronouns and plural words. Note that since the priority of collection is given to structural patterns of CNNCs, the author accepts that the information on non-numeral quantifiers was not particularly focused on. This was due to the limited time frame of the project. As a result, there remain some gaps on this topic. The gaps are marked by a hyphen (-) which means that the information is perhaps available in the source but the issue of non-numeral quantifiers of the language does not receive attention.

5. Answers to Question (8): Noun classes

- *Present*

This answer means that noun class is evident in the language.

- *Absent*

This answer means that noun class is absent in the language.

- *No information*

This answer means that the information about noun class was not available to the thesis author. It might be possible that that feature is not present in the language and therefore that the issue of noun class is not mentioned in the grammar.

Note that in the case that the source of reference is indicated as (*WALS*), this means that the answer to the question was taken from the database made by Corbett (2005). The reader is referred to the *Interactive Reference Tool of the World Atlas of Language Structures* on the feature of *number of genders* for further information such as the actual source of reference.

6. Answers to Question (9): Numeral classifiers

- *Present*

This answer means that numeral classifiers are evident in the language but it is not known clearly whether the system is obligatory or optional.

- *Absent*

This answer means that numeral classifiers are absent in the language.

- *Obligatory*

This answer means that numeral classifiers are present as an obligatory category in the language.

- *Optional*

This answer means that numeral classifiers are present but as an optional category in the language.

Note that in the case that the source of reference is indicated as (*WALS*), this means that the answer to the question was taken from the database made by Gil (2005). The reader is referred to the *Interactive Reference Tool of the World Atlas of Language Structures* on the feature of *numeral classifiers* for further information such as the exact source of reference..

Also, in the case that the source of reference in the survey of reference grammars is indicated as *passim*, this means that the answer comes from the elicitation made from the examples present throughout the grammar and texts. For example, if the answer is “*absent (passim)*”, it means that the elicitation is made

from the observation that the numeral classifier is not mentioned at all in all counting expressions in the language.

Language: Abkhaz
Family/Genus: Northwest Caucasian/Northwest Caucasian
Country/ Macro Area: Georgia/Eurasia
Reference(s): Hewitt, G. (1979) and personal communication*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

ɟyab-k'

girl-one

'one girl' (1979: 57)

2. Restrictions on the patterns in (1) above: *No restrictions (George Hewitt, p.c.)*

3. Structural patterns of CNNC_{NSG}

{N,NUM,NSG}

a-c°a-k ° à

y °ə-bà

ART-apple-PL

two-NONHUM

'two apples' (1979: 121)

{N,NUM}

ʒ°əyza

c°a

eleven

apple

'eleven apples' (1979: 121)

4. Restrictions on the patterns in (3) above

{N,NUM,NSG}: *No known restrictions*

{N,NUM}: *Used with non-human nouns (1979: 121)*

Number systems

5. Number distinction: *SG/PL (1979: 121, passim)*

6. Obligatoriness of number marking: *Obligatory (1979: 149)*

7. Non-numeral quantifiers: *LQ, e.g. rac°a 'many', as in à- ɟyab rac°a [ART-girl many] 'many girls' (1979: 157)*

Classifier systems

8. Noun classes: *Present (1979: 152)*

9. Numeral classifiers: *Absent (WALS)*

* George Hewitt is Professor of Caucasian languages at SOAS, University of London, UK. Email: gh2@soas.ac.uk

Language: Abun
Family/Genus: West Papuan/ North-Central Bird's Head
Country/ Macro Area: Indonesia/Australia-New Guinea
Reference(s): Berry, K. (1995)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{*N, NUM, CLF*}

ndar sye ge at
dog big CLF four

'I found four big dogs.' (1995: 100)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction (WALS)*
6. Obligatoriness of number marking: *See (5)*
7. Non-numeral quantifiers: *LQ, e.g. mwa 'many', as in minda kri (ge) mwa [butterfly yellow (CLF) many] 'many yellow butterflies.'* (1995: 133)

Classifier systems

8. Noun classes: *Absent (1995: passim)*
9. Numeral classifiers: *Obligatory (WALS)*

Language: Acoma
Family/Genus: Keresan/Keresan
Country/ Macro Area: United States(New Mexico)/North America
Reference(s): Miller, W. R. (1965)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

?iska-sái

one-day

'one day' (1965: 166)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

dyâana-sái

four-day

'four days' (1965: 166)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (1965: 166, 178)*

Number systems

5. Number distinction: *SG/DU/PL (1965: 147-148)*

6. Obligatoriness of number marking: *Optional (1965: 148)*

7. Non-numeral quantifiers: *LQ, e.g. háčə 'a few, several', as in háčə mîiya [a few, several mile] 'a few miles, several miles' (1965: 178)*

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Absent (WALS)*

Language: Ainu
Family/Genus: Ainu/Ainu
Country/ Macro Area: Japan/Eurasia
Reference(s): Tamura, S. (2000)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

sine *cise*
one *house*
'one house' (2000: 187)

{N,NUM,CLF}

pon *saro* *sine-p*
be.small *monkey* *one-CLF* (lit. 'thing')
'one small monkey' (2000: 257)

2. Restrictions on the patterns in (1) above:

{N,NUM}: *No known restrictions*

{N,NUM,CLF}: *Used with nouns accompanied by noun modifiers* (2000: 190), hence @.

3. Structural patterns of CNNC_{NSG}

{N,NUM}

tu *cise*
two *house*
'two houses' (2000: 190)

{N,NUM,CLF}

pirka *cise* *tu-p*
be.good *house* *two-CLF* (lit. 'thing')
'two beautiful houses' (2000: 190)

4. Restrictions on the patterns in (3) above: *Same as* (2)

Number systems

5. Number distinction: *No distinction* (2000: 40)

6. Obligatoriness of number marking: *See* (5)

7. Non-numeral quantifiers: *LQ*, e.g. *poronno* 'many', as in *cise poronno* [house many] 'many houses' (2000: 40); *Reduplication*, e.g. *uype* (SG)/*uypeuype* (PL) 'fragment' (2000: 201)

Classifier systems

8. Noun classes: *Absent* (2000: 81)

9. Numeral classifiers: *Optional* (2000: 190)

Language: Akha
Family/Genus: Sino-Tibetan/Burmese-Lolo
Country/ Macro Area: China, Thailand, Myanmar/Southeast Asia & Oceania
Reference(s): Hansson, I-L. (2003) and McDaniel, M. (2002)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

áshì thì shì
fruit one CLF
'one fruit' (2003: 243)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

tshǎhà jǝm̀ xhǝ́ njì yà
person good those two CLF
'Those two good persons' (2003: 241)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction (2003: 242)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: *LQ, e.g. rgawv myah neeh 'many' (2002: 87)*

Classifier systems

8. Noun classes: *Absent (2003: passim)*

9. Numeral classifiers: *Obligatory (2003: 244)*

Language: Albanian
Family/Genus: Indo-European/Albanian
Country/ Macro Area: Serbia and Montenegro, Albania/Eurasia
Reference(s): Camaj, M. (1984), Friedman, V. A. (personal communication)*, Newmark, L. (1957) and Newmark, L., Hubbard, P. and Prifti, P. (1982)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

një burrë

one man

'one man' (1984: 21, 84)

2. Restrictions on the patterns in (1) above: *No restrictions* (Victor A. Friedman, p.c.)

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

dy burra

two man.PL

'two men' (1984: 21, 84)

4. Restrictions on the patterns in (3) above: *No restrictions* (Victor A. Friedman, p.c.)

Number systems

5. Number distinction: *SG/PL* (1982: 139)

6. Obligatoriness of number marking: *Obligatory* (1957: 69)

7. Non-numeral quantifiers: *LQ*, e.g. *disa* 'some', as in *disa fustane kadifeje* [*some dress.PL velvet.ABL*] 'some velvet dresses' (2003: 16)

Classifier systems

8. Noun classes: *Present* (1957: 53-54)

9. Numeral classifiers: *Absent* (WALS)

* Victor A. Friedman is Professor of Slavic languages and literature at the University of Chicago, USA. Email: vfriedm@uchicago.edu

Language: Alutor
Family/Genus: Chukotko-Kamchatkan/Northern Chukotko-Kamchatkan
Country/ Macro Area: Russia (Asia)/Eurasia
Reference(s): Kibrik, A. E., Kodzasov, S. V. and Muravyova, I. A. (2004) and Mel'chuk I. A. (1988).

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,SG}

<i>ənnan</i>	<i>rara-ŋa</i>
<i>one</i>	<i>house-NOM.SG</i>
<i>'one house' (2004: 371, 508)</i>	

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>ŋitaq</i>	<i>ŋavəsqatpi</i>	<i>t[Z]</i>
<i>two</i>	<i>girl</i>	<i>NOM.DU</i>
<i>'two girls' (1988: 275)</i>		

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL (2004: 289)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ, e.g. ʔoptəlʔu 'all' as in ʔoptəlʔu qajunʔu-wwi [all boy-NOM.PL]*
'all boys' (2004: 277)

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Ambulas
Family/Genus: Sepik/Middle Sepik
Country/ Macro Area: Papua New Guinea/Australia-New Guinea
Reference(s): Wilson, P. R. (1980)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:
{*N, NUM*}

<i>néwaa</i>	<i>véti</i>
<i>mother</i>	<i>two</i>

'two mothers' (1980: 116)
4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No information*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (1980: 46-51)*
9. Numeral classifiers: *Absent (WALS)*

Language: Amele
Family/Genus: Trans-New Guinea/Madang
Country/ Macro Area: Papua New Guinea/Australia-New Guinea
Reference(s): Roberts, J. R. (1987)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

dana lecis

man two

'two men' (1987: 157)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (1987: 90, 94)*

Number systems

5. Number distinction: *SG/DU/PL (NB: only inalienably possessed nouns and kinship nouns) (1987: 200)*
6. Obligatoriness of number marking: *Obligatory (1987: 200)*
7. Non-numeral quantifiers: *LQ, e.g. madi 'many' as in dana caja madi [man woman many] 'many people' (1987: 90); 3rd person plural pronoun, e.g. dana (age) ho-ig-a [man (3PL) come-3PL-TOD.PST] 'men came' (1987: 162)*

Classifier systems

8. Noun classes: *Absent (1987: 203)*
9. Numeral classifiers: *Absent (WALS)*

Language: Angolar
Family/Genus: Creoles and Pidgins
Country/ Macro Area: São Tomé e Príncipe/Africa
Reference(s): Lorenzino, G. A. (1998) and personal communication*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

uⁿa ria

one day

'one day' (1998: 137)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

kwin anu

ten year

'ten years' (1998: 135)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction (Augusto Lorenzino, p.c.)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (Augusto Lorenzino, p.c.)*

9. Numeral classifiers: *Absent (Augusto Lorenzino, p.c.)*

* Augusto Lorenzino is Associate Professor of Spanish and Portuguese at Temple University, Philadelphia, USA. Email: galorenz@temple.edu

Language: llAni
Family/Genus: Khoisan/Central Khoisan
Country/ Macro Area: Botswana/Africa
Reference(s): Heine, B (1999) and personal communication*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,SG}

lúí-m' !'úí-mà
 one-3SG.M leopard-3SG.M

'one leopard' (Vossen 1986 in Heine 1999: 38)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

laím kuń xèù- tsà
 two big hippo- M.DU

'two big hippos.' (1999: 34)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: SG/DU/PL (1999: *passim*) (NB: some nouns are transnumeral, e.g. nlnũ 'canoe' or 'canoes' (1999: 30)

6. Obligatoriness of Number marking: *Optional* (Bernd Heine, p.c.)

7. Non-Numeral quantifiers: LQ, e.g. lxóá 'few', as in lxóá xèù [few hippo] 'few hippos' (1999: 39)

Classifier systems

8. Noun classes: *Present* (1999: *passim*)

9. Numeral classifiers: *Absent* (1999: *passim*)

* Bernd Heine is Emeritus Professor of Linguistics at the University of Cologne, Germany. Email: bernd.heine@uni-koeln.de

Language: Arabana
Family/Genus: Australian/ Pama-Nyungan
Country/ Macro Area: Australia (Lake Eyre Basin, South Australia)/Australia-New Guinea
Reference(s): Hercus, L.A. (1994)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

madla *nguyu*

dog *one*

'one dog' (1994: 64)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}

{N,NSG}

putyi-karikari

cat-TRI

'three cats' (1994: 64)

karna-pula

man-DU

'two men' (1994: 65)

{N,NUM}

nhantu *parkulu*

horse *two*

'two horses' (1994: 95)

4. Restrictions on the patterns in (3) above:

{N,NSG}: *Used generally with no known restrictions (NB: trial: used with animate nouns only) (1994: 65, 91); {N,NUM}: No known restrictions*

Number systems

5. Number distinction: *SG/DU/TRI/PL (1994: 63-64)*

6. Obligatoriness of Number marking: *Optional (1994: 65)*

7. Non-Numeral quantifiers: *LQ, e.g. nhuka 'many', as in wardukupa nhuka [many child] 'many children' (1994: 95)*

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Absent (1994: passim)*

Language: Arabic (Modern Standard)
Family/Genus: Afro-Asiatic/Semitic
Country/ Macro Area: Saudi Arabia and adjacent countries in Middle East and North Africa/Africa
Reference(s): Holes, C. (1995), (2004) and personal communication¹ and Kaye, A. S. (1987)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

kita:b *wa:ħid*
book *one*
 'one book' (1995: 173)

2. Restrictions on the patterns in (1) above: *No restrictions (Clive Holes, p.c.)*

3. Structural patterns of CNNC_{NSG}

{N,NUM,NSG}

kita:b-a:ni *θna:ni*
book-DU *two.M*
 'two books' (1995: 173)

{N,NUM,OBL,NSG}

ʔarba 'at-u *ʔawla:d-in*
four.F-NOM *boy.PL-GEN*
 'four boys' (1995: 173)

{N,NUM,ACC,SG}

xams-a *'ašarat-a* *bint-an*
five.M-ACC *ten.F-ACC* *girl-ACC*
 'fifteen girls' (lit. 'fifteen as regards girl') (1995: 174)

{N,NUM,OBL,SG}

mi ʔat-u *bint-in*
hundred-NOM *girl-GEN*
 'a hundred girls' (1995: 175)

4. Restrictions on the patterns in (3) above

{N,NUM,NSG} (NSG=DU): *Used with the numeral 2 (1995: 173)*

{N,NUM,OBL,NSG}: *Used with the numerals 3-10 (1995: 173)*

{N,NUM,ACC,SG}: *Used with the numerals 11-99 (1995: 174)*

{N,NUM,OBL,SG}: *Used with the high round numerals, e.g. 100 (1995: 175)*

* Clive Holes is Professor for the Study of the Contemporary Arab World at the University of Oxford, UK. Email: clive.holes@orinst.ox.ac.uk

Number systems

5. Number distinction: *SG/DU/Pl* (1995: 173)

6. Obligatoriness of Number marking: *Obligatory* (1987: 678)

7. Non-Numeral quantifiers: *LQ*, e.g. *kaθi:r* 'many', as in *l-kaθi:r mina n-na:s* [*DEF-many of the-people*] 'many people' (1995: 162)

Classifier systems

8. Noun classes: *Present* (2004: 174)

9. Numeral classifiers: *Absent* (*Clive Holes*, p.c.)

Language: Arapesh
Family/Genus: Torricelli/Kombio-Arapesh
Country/ Macro Area: Papua New Guinea/Australia-New Guinea
Reference(s): Conrad, R. J. and Wogiga, K. (1991)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:
{*N,NUM,NSG* }
bwi-yogw *beto-gw*
two-PL.CL *bed-PL*
'two beds' (1991: 59)
4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (WALS)*
6. Obligatoriness of Number marking: *No information*
7. Non-Numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (WALS)*
9. Numeral classifiers: *No information*

Language: Armenian (Eastern)
Family/Genus: Indo-European/Armenian
Country/ Macro Area: Armenia/Eurasia
Reference(s): Campbell, G. L. (2000), Kozintseva, N. (1995) and Stevick, E. W. (1955)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

erek' *usanoԿ*

three *student*

'three students' (1995: 8)

{N,NUM,CLF}

yerk'u *hoki* *zinvor*

two *soul (lit. 'soul')* *soldier*

'two soldiers' (1955: 30) (soul = numeral classifier-like word)

{N,NUM,CLF,NSG}

mi *vic* *hoki* *k'anayk*

INDEF six *CLF (lit. soul)* *woman.PL*

'some six women' (1955: 30)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2000: 124)*
6. Obligatoriness of Number marking: *Obligatory (WALS)*
7. Non-Numeral quantifiers: *LQ, e.g. bolor 'all', as in bolor erexa-ner-ə [all child-PL-DEF] 'all children' (1995: 12)*

Classifier systems

8. Noun classes: *Absent (2000: 124)*
9. Numeral classifiers: *Optional (1955: passim)*

Language: Armenian (Western)
Family/Genus: Indo-European/Armenian
Country/ Macro Area: Turkey/Eurasia
Reference(s): Campbell, G. L. (2000) and Sigler, M. (1990)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>ksan</i>	<i>usanoY</i>
<i>twenty</i>	<i>student</i>
<i>'twenty students' (1990: 500)</i>	

{N,NUM,NSG}

<i>ksan</i>	<i>usanoY-ner</i>
<i>twenty</i>	<i>student-PL</i>
<i>'twenty students' (1990: 500)</i>	

4. Restrictions on the patterns in (3) above:

{N,NUM}: When *N* is the subject "may trigger plural marking on verb or not" (1990: 500)

{N,NUM,NSG}: When *N* is the subject "must trigger plural verbal agreement" (1990: .500)

Number systems

5. Number distinction: *SG/PL (1990: passim)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (2000: 124)*
9. Numeral classifiers: *No information*

Language: Atayal
Family/Genus: Austronesian/Atayalic
Country/ Macro Area: Taiwan /Southeast Asia & Oceania
Reference(s): Rau, D-H. V. (1992) and Philips Davis (personal communication)*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

<i>qutux</i>	<i>squliq</i>
<i>one</i>	<i>person</i>
<i>'one person'</i>	<i>(1992: 131)</i>

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>saziŋ</i>	<i>yonomi</i>
<i>two</i>	<i>cup</i>
<i>'two cups'</i>	<i>(1992: 131)</i>

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No information*

6. Obligatoriness of Number marking: *No information*

7. Non-Numeral quantifiers: *Reduplication, e.g. btunux (SG)/b-btunux (PL) 'stone' (1992: 116)*

Classifier systems

8. Noun classes: *Absent (Philips Davis, p.c.)*

9. Numeral classifiers: *Absent (WALS)*

* Philips Davis is Emeritus Professor of Linguistics at Rice University, USA. Email: pwd@rice.edu

Language: Awa Pit
Family/Genus: Barbacoan/Barbacoan
Country/ Macro Area: Colombia,Ecuador/South America
Reference(s): Curnow, T. J. (1997)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

maza kuzhu

one pig

'one pig.' (1997: 75)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

paas paynkul

two son

'two sons' (1997: 77)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (NB: verbal suffixes denoting plurality) (1997: 66, 75)*

6. Obligatoriness of Number marking: *Optional (1997: 66, 185)*

7. Non-Numeral quantifiers: *LQ, e.g. akkwan 'many', as in akkwan awa [many person] 'many people' (1997: 184)*

Classifier systems

8. Noun classes: *Absent (1997: 83)*

9. Numeral classifiers: *Absent (1997: passim)*

Language: Awtuw
Family/Genus: Sepik/Ram
Country/ Macro Area: Papua New Guinea /Australia-New Guinea
Reference(s): Feldman, H. (1986) and personal communication*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

yæn naydowo

child one

'one child' (1986: 123)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

yæn-wæw yikiyr

child-DU two

'two children' (1986: 123)

{N,NUM}

tale yikiyr

woman two

'two women' (1986: 123)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL* (1986: 40)

6. Obligatoriness of Number marking: *Optional* (1986: 40)

7. Non-Numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent* (Harry Feldman, p.c.)

9. Numeral classifiers: *Absent* (WALS)

* Harry Feldman is a linguist at Quaid-Azam University, Islamabad, Pakistan. Email: harry.feldman@gmail.com

Language: Bai
Family/Genus: Sino-Tibetan/Bai
Country/ Macro Area: China /Southeast Asia
Reference(s): Wiersma, G. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:
{N,NUM,CLF}

<i>sʋ̂</i>	<i>ŋʋ</i>	<i>tsʰuε²</i>
<i>book</i>	<i>five</i>	<i>CLF</i>

'five books' (p.671)
4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No information*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (2003: passim)*
9. Numeral classifiers: *Present (2003: passim)*

Language: Barasano
Family/Genus: Tucanoan/Tucanoan
Country/ Macro Area: Colombia/South America
Reference(s): Jones, W. and Jones, P. (1991)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}

{N,NUM,CLF,NSG}

hʉa-rahe *kōbe-rahe-ri*

two-CLF *metal-CLF-PL*

'two metal cans' (1991: 59) (NB: *rahe* 'cylinder', used as numeral classifier)

{N,NUM,CLF}

gũta-hāī *hʉa-hāī*

stone-CLF *two-CLF*

'two stone slabs' (1991: 59)

4. Restrictions on the patterns in (3) above *No known restrictions*

Number systems

5. Number distinction: *SG/PL (WALS)*
6. Obligatoriness of number marking: *Obligatory (WALS)(NB: the example in (3))*
7. Non-numeral quantifiers: *LQ, e.g. hārā 'many', as in hārā bās-a [many human-PL] 'many humans' (1991: 59)*

Classifier systems

8. Noun classes: *Present (WALS)*
9. Numeral classifiers: *Present (WALS)*

Language: Basque
Family/Genus: Basque/Basque
Country/ Macro Area: France, Spain /Eurasia
Reference(s): Hualde, J. I. and Urbina, J.O. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

liburu bat

book one

'one book' (2003: 136)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (2003: 136, 386)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

bi etxe

two house

'two houses' (2003: 136)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (2003: 136)*

Number systems

5. Number distinction: *SG/PL (2003: 432)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: *LQ, e.g. asko 'lots of', as in gutun asko [letter a.lot.ABS] 'a lot of letters' (2003: 431)*

Classifier systems

8. Noun classes: *Absent (2003: 113)*

9. Numeral classifiers: *Absent (2003: 113)*

Language: Bauzi
Family/Genus: East Geelvink Bay/ East Geelvink Bay
Country/ Macro Area: Indonesia /Australia-New Guinea
Reference(s): Briley, D. (1997)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

nao *vàmtea*

leg *one*

'one leg' (1997: 17)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

dam *behàsu*

man.PL *two*

'two men' (1997: 89)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No information*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Begak (Ida'an)
Family/Genus: Austronesian/Borneo
Country/ Macro Area: Malaysia/Southeast Asia & Oceania
Reference(s): Goudswaard, N. (2005)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

sə-tassa' asu

one-CLF dog

'one dog' (2005: 101)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

duo tassa' asu

two CLF dog

'two dogs' (2005: 89)

{N,NUM}

duo anak

two child

'two children' (2005: 272)

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF}: *Used generally with no known restrictions (2005: 273)*

{N,NUM}: *Rare (2005: 273), hence @*

Number systems

5. Number distinction: *No distinction (2005: passim)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: *Reduplication, e.g. anak (SG)/anak-anak (PL) 'child' (2005: 276)*

Classifier systems

8. Noun classes: *Absent (2005: passim)*

9. Numeral classifiers: *Optional (2005: 102, passim)*

Language: Bengali
Family/Genus: Indo-European/Indic
Country/ Macro Area: Bangladesh, India/Eurasia
Reference(s): Dasgupta, P. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

æktā *boi*

one.CLF *book*

'one book' (2003: 380)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (2003: 366)*

3. Structural patterns of CNNC_{NSG}:

{N,NUMCLF}

duo *boi*

two.CLF *book*

'two books' (2003: 380)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (2003: 366, 380)*

Number systems

5. Number distinction: *SG/PL (2003: 365)*

6. Obligatoriness of number marking: *Obligatory (only animate) (2003: 365)*

7. Non-numeral quantifiers: *LQ, e.g. anek 'many', as in anek mohilā [many woman] 'many women' (2003: 366)*

Classifier systems

8. Noun classes: *Absent (2003: 367)*

9. Numeral classifiers: *Obligatory (WALS)*

Language: Berber (Ayt Seghrouchen Middle Atlas)
Family/Genus: Afro-Asiatic/Berber
Country/ Macro Area: Morocco /Africa
Reference(s): Penchoen, T. G. (1973)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,SG}

yun *u-ryaz*
one *SG-man*

'one man' (1973: 25) (see also 1973:14, 41 for notes on number)

{N,NUM,OBL,SG}

yut *n-t-məttutt*
one *of-SG-woman*

'one woman' (1973: 25) (see also 1973:14, 41 for notes on number)

2. Restrictions on the patterns in (1) above:

{N,NUM,SG}: Used with nouns beginning with a vowel (i.e. masculine nouns)(1973: 25)

{N,NUM,OBL,SG}: Used with nouns beginning with a consonant (i.e. feminine nouns and unberberized Arabic loans) (1973: 25)

3. Structural patterns of CNNC_{NSG}

{N,NUM,NSG}

sin *i-rəyz-ən*
two *PL-man-PL*

'two men' (1973: 25)

{N,NUM,SG}

εəʃrin *u-ryaz*
twenty *SG-man*

'twenty man' (1973: 25)

{N,NUM,OBL,NSG}

snat *n-təwtm-in*
two *of-woman-PL*

'two woman' (1973: 25)

{N,NUM,OBL,SG}

εəʃrin *n-t-məttutt*
twenty *of-SG-woman*

'twenty women' (1973: 25)

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG}: Used with masculine nouns modified by the numerals 2-10 (1973: 25)

{N,NUM,SG}: Used with masculine nouns modified by the numerals above 10 (1973: 25)

{N,NUM,OBL,NSG}: Used with feminine nouns and unberberized nouns modified by the numerals 2-10 (1973: 25)

{N,NUM,OBL,SG}: Used with feminine nouns and unberberized nouns modified by the numerals above 10 (1973: 25)

Number systems

5. Number distinction: *SG/PL* (1973: 14)

6. Obligatoriness of number marking: *Obligatory* (1973: 14)

7. Non-numeral quantifiers: *LQ*, e.g. *ša* 'some', as in *ša i-rəyz-ən* [*some PL-man-PL*] 'some men' (1973: 60)

Classifier systems

8. Noun classes: *Present* (1973: 25, *passim*)

9. Numeral classifiers: *Absent* (1973: *passim*)

Language: Berbice Dutch Creole
Family/Genus: Creoles and Pidgins
Country/ Macro Area: Guyana/South America
Reference(s): Kouwenberg, S. (1994)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

en jɛrma

one woman

'one woman' (1994: 152)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (1994: 50, 152, 154)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

twe jɛrmatoko

two woman.child

'two daughters' (1994: 239)

{N,NUM,NSG}

twe sosɔr-apu

two sister-PL

'two sisters' (1994: 239)

4. Restrictions on the patterns in (3) above:

{N,NUM}: *No restrictions (1994: 239)*

{N,NUM,NSG}: *Exceptional (1994: 239), hence @*

Number systems

5. Number distinction: *SG/PL (1994: 238)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ, e.g. alma 'all', as in almek aka [all.ISG tooth] 'all my teeth' (1994: 95)*

Classifier systems

8. Noun classes: *Absent (1994: 61)*

9. Numeral classifiers: *Absent (1994: passim)*

Language: Berta
Family/Genus: Nilo-Saharan/Berta
Country/ Macro Area: Ethiopia,Sudan /Africa
Reference(s): Triulzi (1976) and Tucker, A.N. and Bryan, M.A. (1966)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

mis' ε hohulu dukunu

bird egg one

'one bird's egg' (1966: 351)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

mabi hoolon

man.PL two

'two men' (1966: 351) (NB: maba = SG, 1966: 349)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1976: 524)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (1966: 349)*

9. Numeral classifiers: *No information*

Language: Bhumij
Family/Genus: Austro-Asiatic/Munda
Country/ Macro Area: India/Eurasia
Reference(s): Ramaswami, N. (1992)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

moyon bɔi

one book

'one book' (1992: 85)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

baria sadɔm

two horse

'two horses'. (1992: 95)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL* (1992: 6,75)

6. Obligatoriness of number marking: *Obligatory* (1992: 6)

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present* (NB: *not marked consistently*) (1992: 73)

9. Numeral classifiers: *No information*

Language: Binandere
Family/Genus: Trans-New Guinea/ Binanderean
Country/ Macro Area: Papua New Guinea/Australia-New Guinea
Reference(s): Wilson, J.P. (1996) and (2002)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

eutu *da*
woman *one*
 'one woman' (1996: 58)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

ma-mai *tote*
PL-child *two*
 'two children' (1996: 13)

{N,NUM}

kasiwo *babain* *tote*
big.knife *big* *two*
 'two big knives' (1996: 87)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

- 5. Number distinction: *No information*
- 6. Obligatoriness of number marking: *No information*
- 7. Non-numeral quantifiers: *LQ*, e.g. *jiwae* 'many', as in *bido jiwae* [*many banana*] 'many bananas' (2002: 29); *Reduplication*, e.g. *mamai* [*child.REDUP*] 'children' (2002: 12)

Classifier systems

- 8. Noun classes: *No information*
- 9. Numeral classifiers: *No information*

Language: Bislama
Family/Genus: Creoles and Pidgins
Country/ Macro Area: Vanuatu/Southeast Asia & Oceania
Reference(s): Crowley, T. (2004) and Miriam Meyerhoff (personal communication)*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

wan man

one man

'one man' (2004: 26)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions* (2004: 26)

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

sikis man

six man

'six people' (2004: 29)

{N,NUM,CLF}

tu frut sigaret

two CLF cigarette

'two cigarettes' (2004: 56)

4. Restrictions on the patterns in (3) above:

{N,NUM}: *Used generally with no known restrictions* (2004: 29, 56)

{N,NUM,CLF}: *Used with a few nouns, hence @* (2004: 56)

Number systems

5. Number distinction: *SG/PL* (2004: *passim*)

6. Obligatoriness of number marking: *Obligatory (but not with overt number)* (Miriam Meyerhoff, *p.c.*)

7. Non-numeral quantifiers: *LQ*, e.g. *plante* 'many', as in *plante woman* [*plenty woman*] 'plenty of women' (2004: 37)

Classifier systems

8. Noun classes: *Absent* (Miriam Meyerhoff, *p.c.*)

9. Numeral classifiers: *Absent* (2004: *passim*); (*NB: a few nouns require a numeral classifier-like word*)

* Miriam Meyerhoff is Professor of Sociolinguistics at the University of Edinburgh, UK. Email: Miriam.Meyerhoff@ed.ac.uk

Language: Bribri
Family/Genus: Chibchan/Talamanca
Country/ Macro Area: Costa Rica /North America
Reference(s): Campbell, G. (2000) and Yasugi, Y. (1995)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>Je</i>	<i>kj</i>	<i>alà-r</i>	<i>kianq-dak</i>	<i>bul</i>
<i>1SG</i>	<i>EXP</i>	<i>child-PL</i>	<i>want.PST-PL</i>	<i>two</i>

'I wanted two children.' (Dickeman Datz 1984: 116 in Yasugi 1995: 403)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No information*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (2000: 264)*
9. Numeral classifiers: *No information*

Language: Broken
Family/Genus: Creoles and Pidgins
Country/ Macro Area: Australia (Queensland)/Australia-New Guinea
Reference(s): Shnukal, A. (1988)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

wan seya

one chair

'one chair' (1988: 28)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (1988: 25, 28)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

tu an

two hand

'two hands' (1988: 28)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (1988: 28)*

Number systems

5. Number distinction: *No distinction (1988: 23)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: *LQ, e.g. plenti 'a lot of', as in plenti pikinini [a.lot.of child] 'a lot of children' (1988: 29); Definite article, e.g. dem gel [DEF.ART.PL girl] 'the girls' (1988: 24)*

Classifier systems

8. Noun classes: *Absent (1988: 23)*

9. Numeral classifiers: *Absent (1988: passim)*

Language: Buglere
Family/Genus: Chibchan/Guaymi
Country/ Macro Area: Panama/South America
Reference(s): Yasugi, Y. (1995)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

glí *gadá-de*

tree *CLF-one*

'one tree' (Solis Hernandez 1989: 149-150 in Yasugi 1995: 405)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (Solis Hernandez 1989: 149-150 in Yasugi 1995: 405)*

3. Structural patterns of CNNC_{NSG}: *No information*

4. Restrictions on the patterns in (3) above: *No information*

Number systems

5. Number distinction: *No information*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Present (Solis Hernandez 1989: 149-150 in Yasugi 1995: 405)*

Language: Bunuba
Family/Genus: Australian/Bunuban
Country/ Macro Area: Australia (Western Australia)/Australia-New Guinea
Reference(s): Rumsey, A. (2000)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

gawarra *yuwana*

sun/hour *one*

'one hour/one o'clock' (2000: 113)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}

{N,NSG}

buga-arri

child-DU

'two children' (2000: 65)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (2000: 65, passim)*

Number systems

5. Number distinction: *SG/DU/PL (2000: 64, 115)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *Reduplication, e.g. ngarranayi (SG)/ngarrngarranyi (PL) 'mother' (2000: 69)*

Classifier systems

8. Noun classes: *Absent (2000: passim)*

9. Numeral classifiers: *Absent (WALS)*

Language: Buru
Family/Genus: Austronesian/Central Malayo-Polynesian
Country/ Macro Area: Indonesia/Southeast Asia & Oceania
Reference(s): Grimes, C. E. (1991)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

huma *emsian*
house *one*
'one house' (1991: 294)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

huma-r *polo*
house-PL *ten*
'ten houses' (1991: 295)

{N,NUM,(CLF)}

fafu (*kisen*) *rua*
pig (CLF) *two*
'two pigs' (1991: 307)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1991: 295)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ, e.g. edemen 'many', as in fafu edemen [pig many] 'many pigs' (1991)*

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Optional (1991: 307)*

Language: Burushaski
Family/Genus: Burushaski/Burushaski
Country/ Macro Area: Pakistan / Eurasia
Reference(s): Lorimer, D. (1935)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

hik bΛda

one pace,step

'one pace,step' (1935: 191)

{N,SG}

gərk-ən

pea(PL)-SG

'a pea' (1935: 48)

{N,NUM,SG}

hΛn huyεsΛn

one goat/sheep(PL).SG

'a goat/sheep' (1935: 48)

2. Restrictions on the patterns in (1) above:

{N,NUM}: Used generally with no known restrictions (1935: 191, *passim*)

{N,SG} and *{N,NUM,SG}*: Used with collective nouns (1935: 46-48, *passim*), hence @

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

alta huyεs

two goat(PL)

'two goats' (1935: 48)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1935: 26)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (1935: 14)*

9. Numeral classifiers: *Absent (WALS)*

Language: Camling
Family/Genus: Sino-Tibetan/Bodic
Country/ Macro Area: Nepal/Southeast Asia & Oceania
Reference(s): Ebert, K. H. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

i-li suntala

one-CLF orange

'one organge' (2003: 536)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF,NSG}

haka-po maricha-ci

two-CLF girl-PL

'two girls' (2003: 536)

{N,NUM,CLF}

sim-ra pyupa

three-CLF cow

'three cows' (2003: 536)

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF,NSG}: Used with human referents (2003: 536)

{N,NUM,CLF}: Used with non-human referents (2003: 536)

Number systems

5. Number distinction: *SG/PL (NB: only human nouns) (2003: 536)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Present (only with numerals 1-3) (2003: 536)*

Language: Carrier
Family/Genus: Na-Dene/Athapaskan
Country/ Macro Area: Canada/North America
Reference(s): Green, J. (2006) and Mithun, M. (1999)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF,NSG}

<i>tane</i>	<i>lhi-ke</i>
<i>three.CLF</i>	<i>dog-PL</i>
<i>'three dogs' (2006: 1)</i>	

{N,NUM,CLF}

<i>ndən</i>	<i>yoh</i>
<i>two.CLF</i>	<i>house</i>
<i>'two houses' (1999: 104)</i>	

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF}: *No known restrictions*

{N,NUM,CLF,NSG}: *Used with nouns denoting humans and dogs (2006: 1)*

Number systems

5. Number distinction: *SG/PL (NB: only nouns denoting humans and dogs) (2006: 1)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: *LQ, e.g. lhadun yoh [many.CLF house] 'many houses' (2006: 1)*

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *Obligatory (WALS)*

Language: Catalan
Family/Genus: Indo-European/Romance
Country/ Macro Area: Spain/Eurasia
Reference(s): Hualde, J. I. (1992)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

<i>un</i>	<i>noi</i>
<i>one.M</i>	<i>boy (M)</i>
<i>'one boy' (1992: 122)</i>	

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (1992: 122, 127)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>dos</i>	<i>noi-s</i>
<i>two.M</i>	<i>boy (M)-PL</i>
<i>'two boys' (1992: 122)</i>	

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (1992: 122, 126, 127)*

Number systems

5. Number distinction: *SG/PL (1992: passim)*

6. Obligatoriness of number marking: *Obligatory (1992: 100, 116)*

7. Non-numeral quantifiers: *LQ, e.g. algun 'some', as in alguns amics [some-PL friend-PL] 'some friends' (1992: 121)*

Classifier systems

8. Noun classes: *Present (1992: 100)*

9. Numeral classifiers: *Absent (1992: passim)*

Language: Cavineña
Family/Genus: Tacanan/Tacanan
Country/ Macro Area: Bolivia/South America
Reference(s): Guillaume, A. (2004)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

peadya *ekwita*

one *person*

'one person' (2004: 510)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

beta *kwaba*

two *canoe*

'two canoes'

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL (2004: 497, passim)*

6. Obligatoriness of number marking: *Optional (2004: 497-498)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Cayuvava
Family/Genus: Cayuvava/Cayuvava
Country/ Macro Area: Bolivia/South America
Reference(s): Key, H. H. (1967)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

karata nāræ

one tree

'one tree' (1967: 45, 59)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

mitia tōrene

two woman

'two women' (1967: 49-50)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1967: 50)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ, e.g. padetahi 'many', as in padetahi pæñæ [many time] 'many times' (1967: 61)*

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Absent (WALS)*

Language: Chaha
Family/Genus: Afro-Asiatic/Semitic
Country/ Macro Area: Ethiopia/Africa
Reference(s): Ojeda, A. E. (1994)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

at *bet*
one *house*
'one house' (1994: 1)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

xwet *bet*
two *house*
'two houses' (1994: 1)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction (1994: 1)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Chantyal
Family/Genus: Sino-Tibetan/Bodic
Country/ Macro Area: Nepal/Southeast Asia & Oceania
Reference(s): Noonan, M. (1999), (2003) and personal communication*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

yəw-ta *pɦale*

one-CLF *leg*

'one leg' (1999: 551)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

dwi-ta *pɦale*

two *leg*

'two legs' (1999: 557)

{N,NUM,CLF,NSG}

tin-ta *jəmməy* *naku-ma*

three-CLF *all* *dog-PL*

'all three dogs' (2003: 318)

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF}: *Used generally with no known restrictions (1999: 557, 566)*

{N,NUM,CLF,NSG}: *Rare (Michael Noonan, p.c.), hence @*

Number systems

5. Number distinction: *SG/PL (2003: 318)*

6. Obligatoriness of number marking: *Optional (2003: 318)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Optional (2003: 321)*

* Michael Noonan is Professor of Linguistics at the University of Wisconsin, USA. Email: noonan@csd.uwm.edu

Language: Chimila
Family/Genus: Chimila/Chimila
Country/ Macro Area: Colombia /South America
Reference(s): Adelaar, W. and Muysken, P. (2004) and Willem Adelaar (personal communication)*

CNNCs

- 1. Structural patterns of CNNC_{SG}: *No information*
- 2. Restrictions on the patterns in (1) above: *No information*
- 3. Structural patterns of CNNC_{NSG}:

{N,NUM}

na^{ʔn} bu^hna ogg^we

1SG.POSS two child

'my two children' (Trillos Amya 1997: 141 in 2004)

- 4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

- 5. Number distinction: *SG/DU/PL (Willem Adelaar, p.c.)*
- 6. Obligatoriness of number marking: *No information*
- 7. Non-numeral quantifiers: -

Classifier systems

- 8. Noun classes: *Absent (Willem Adelaar, p.c.)*
- 9. Numeral classifiers: *Present (Trillos Amya 1997: 139 in 2004)*

* Willem Adelaar is Professor of Native American Languages and Cultures at Leiden University, the Netherlands. Email: W.F.H.Adelaar@let.leidenuniv.nl

Language: Chinantec (Comaltepec)
Family/Genus: Oto-Manguean/Chinantecan
Country/ Macro Area: Mexico/North America
Reference(s): Anderson, J. L. (1989)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

ha:n^L hóʔ^L
one animal
'a/one animal' (1989: 61)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

gi^L-tu^{LM} hi^L
ten-two book
'twelve books' (1989: 58)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction (1989: 61)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: *Plural word, e.g. ha^Lléʔ^L hi^L [PL book] 'books' (1989: 61)*

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Absent (NB: numeral classifier-like words for some (mass) nouns, e.g. tu^M*

ʔni^{LM} mi^H ma^Lhã^{LH}r [two three seed tooth.PL] 'two or three of his teeth' (1989: 59)

Language: Chipewyan
Family/Genus: Na-Dene/Athapaskan
Country/ Macro Area: Canada /North America
Reference(s): Cook, E. (2004)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

ʔiʔághɪ *ts'ákuɪ*
 one *old.woman*
 '*an old woman*' (2004: 103)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

náke *dzirita*
 two *week*
 '*two weeks*' (2004: 100)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL* (2004: 114) (NB: *number marked on verb—verb inflection 2004: 118*)

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language:	Chitimacha
Family/Genus:	Chitimacha/Chitimacha
Country/ Macro Area:	United States (Louisiana)/North America
Reference(s):	Granberry, J. (2004) and Nichols, J. (1992)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

qupa pānx

two man

'two men' (2004: 97)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL only kinship terms, e.g. gimnix (SG)/ginkgank(PL) 'daughter' (NB: most nouns are neutral in regards to number, e.g. hānā, 'house' or 'houses' (2004: 79)*
6. Obligatoriness of number marking: *Obligatory (WALS)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (1992: 301)*
9. Numeral classifiers: *Absent (WALS)*

Language: Chontal Maya
Family/Genus: Mayan/Mayan
Country/ Macro Area: Mexico/North America
Reference(s): Knowles, S. M. (1984) and Nichols, J. (1992)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

ʔun tu noh toro
one CLF large bull
'one large bull' (1984: 239)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

čäm-p'e ʔotot
four-CLF house
'four houses' (1984: 204)

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF}: *Used with the numerals 1-6 which are Chontal numerals; numeral classifiers are not used with the Spanish numerals.*

Number systems

5. Number distinction: *SG/PL (1984: 203, passim)*

6. Obligatoriness of number marking: *Optional (1984: 203)*

7. Non-numeral quantifiers: *LQ, e.g. k'en 'much', as in mäh k'en wah [very much tortilla] 'lots of tortillas' (1984: 204)*

Classifier systems

8. Noun classes: *Absent (1992: 301)*

9. Numeral classifiers: *Present (1984: passim)*

Language: Chrau
Family/Genus: Austro-Asiatic/Bahnaric
Country/ Macro Area: Vietnam/Southeast Asia & Oceania
Reference(s): Thomas, D. D. (1971)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

<i>du</i>	<i>vanông</i>	<i>gapu</i>
<i>one</i>	<i>CLF</i>	<i>buffalo</i>

'one buffalo' (1971: 133)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}: *No information*

4. Restrictions on the patterns in (3) above: *No information*

Number systems

5. Number distinction: *No distinction (WALS)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Present (1971: 127)*

Language: Cree (Plains)
Family/Genus: Algic/Algonquian
Country/ Macro Area: Canada/North America
Reference(s): Wolfart, H. C. (1981)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

pēyak *sīsīp*

one *duck*

'one duck' (1981: 57)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}

{N,NUM,NSG}

nīso *sīsīp-ak*

two *duck-3PL*

'two ducks' (1981: 57)

4. Restrictions on the patterns in (3) above *No known restrictions*

Number systems

5. Number distinction: *SG/PL* (1981: 15-17)

6. Obligatoriness of number marking: *Obligatory* (WALS)

7. Non-numeral quantifiers: *LQ*, e.g. *mihcētiwa* 'be.numerous', as in *mōhkomāna mihcētiwa* [*knife be.numerous*] 'There were many knives' (1981: 55)

Classifier systems

8. Noun classes: *Present* (1981: 19)

9. Numeral classifiers: *Absent* (WALS)

Language: Degema
Family/Genus: Niger-Congo/Edoi
Country/ Macro Area: Nigeria /Africa
Reference(s): Kari, E.E. (1997) and (2004)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,SG}

<i>ó-mó</i>	<i>ó-vu</i>
<i>SG-child</i>	<i>SG-one</i>
<i>'one child' (2004: 239)</i>	

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (2004: 239; 1997: 27)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>iyów</i>	<i>átuuw</i>
<i>twenty</i>	<i>hat.PL</i>
<i>'twenty hats' (1997: 36)</i>	

{N,NUM,NSG,DPM}

<i>ímó</i>	<i>mé</i>	<i>va</i>
<i>child.PL</i>	<i>DPM</i>	<i>two</i>
<i>'two children' (2004: 209)</i>		

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG}: *Used generally with no known restrictions (2004: 209; 1997: 27)*

{N,NUM,NSG,DPM}: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2004: 240)*

6. Obligatoriness of number marking: *Obligatory (2004: .238)*

7. Non-numeral quantifiers: *LQ, e.g. ívu 'some', as in inám ívu [animal.PL some] 'some animals' (1997: 40)*

Classifier systems

8. Noun classes: *Present (2004: 243)*

9. Numeral classifiers: *Absent (2004: passim)*

Language: Diyari
Family/Genus: Australian/Pama-Nyungan
Country/ Macro Area: Australia (South Australia)/Australia-New Guinea
Reference(s): Austin, P. (1981), and personal communication*

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}

{N,NSG}

mankada-wuła

girl-DU

'two girls' (1981: 128)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL (1981: passim)*
6. Obligatoriness of number marking: *Optional (Peter Austin, p.c.)*
7. Non-numeral quantifiers: *LQ, e.g. marapu 'many', as in diṭi marapu [day many] 'many days' (1981: 103); Reduplication, e.g. kupa 'child'/ kupa-kupa 'children' (p.105); Pronoun, e.g. pula '3DU', e.g. pula kaku-yali wapa-yi [3DU.SBJ elder sister-ERG go-PRES] 'The two sisters are going.' (1981: 122)*

Classifier systems

8. Noun classes: *Present (WALS)*
9. Numeral classifiers: *Absent (Peter Austin, p.c.)*

* Peter Austin is Professor of Linguistics at SOAS, University of London, UK. Email: pa2@soas.ac.uk

Language:	Dogon
Family/Genus:	Niger-Congo/Dogon
Country/ Macro Area:	Burkina Faso, Mali/Africa
Reference(s):	Plungian, V. (1995) and personal communication*, and Williamson, K. and Blench, R. (2000)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:
{*N,NUM*}
somu pelu
ten horse
'ten horses' (1995:11)
4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (1995: 11)*

Number systems

5. Number distinction: *SG/PL (WALS)*
6. Obligatoriness of number marking: *Optional (NB: perhaps obligatory for some classes of nouns) (Vladimir Plungian, p.c.)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (but remnant) (2000: 24)*
9. Numeral classifiers: *Absent (Vladimir Plungian, p.c.)*

* Vladimir Plungian is Professor of Linguistics at Institute of Linguistics, Russian Academy of Science, Moscow, Russia. Email: plungian@gmail.com

Language: Dulong
Family/Genus: Sino-Tibetan/Nungish
Country/ Macro Area: China (Tibet)/Southeast Asia & Oceania
Reference(s): LaPolla, R. J. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:
{N,NUM,CLF}
zǎjè ǎjī pǎŋ
book two CLF
'two books' (2003: 676)
4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (2003: 680)*

Number systems

5. Number distinction: *SG/PL (2003: 675-676)*
6. Obligatoriness of number marking: *Optional (2003: 676, 681)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *Present (2003: 675, passim)*

Language: Dumo
Family/Genus: Sko/Western Sko
Country/ Macro Area: Papua New Guinea/Australia-New Guinea
Reference(s): Ross, M. (1980)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

dij yumonu

bird two

'two birds' (1980: 84)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction*
6. Obligatoriness of number marking: *See (5)*
7. Non-numeral quantifiers: *LQ, e.g. podí 'many', as in da plápeŋ podí [pig black many] '(the) black pigs'*

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *Absent (WALS)*

Language: Ejagham
Family/Genus: Niger-Congo/Bantoid
Country/ Macro Area: Cameroon,Nigeria/Africa
Reference(s): Watters, J.R. (1981)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,SG}

N-díg *má-d*

CL-rope *CL-one*

'one rope' (1981: 469) (Class *N* = singular number)

{N,NUM,CLF,SG}

è-rám *í-čókùd* *já-d*

CL-CLF *GEN¹* *CL-orange* *CL-one*

'one orange' (1981: 469) (Class *è-* = singular number)

2. Restrictions on the patterns in (1) above:

{N,NUM,SG}: Used generally with no known restrictions (1981: passim)

{N,NUM,CLF,SG}: Used only with some nouns denoting plants or trees (1981: 313), hence @

3. Structural patterns of CNNC_{NSG}: No information

4. Restrictions on the patterns in (3) above: No information

Number systems

5. Number distinction: SG/PL (1981: passim)

6. Obligatoriness of number marking: Obligatory (1981: passim)

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: Present (1981: passim)

9. Numeral classifiers: Present (NB: not productive) (1981: 313)

¹ Aikhenvald (2000: 99) refers to the tone as 'genitive linker'.

Language: English
Family/Genus: Indo-European/Germanic
Country/ Macro Area: United Kingdom/Eurasia
Reference(s): Jennifer Sullivan (personal communication)*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

wan bɔi

one boy

'one boy'

2. Restrictions on the patterns in (1) above: *No restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

tu: bɔiz

two boy-PL

'two boys'

{N,NUM}

tu: ʃi:p

two sheep

'two sheep'

{N,NUM,OBL,CLF}

tu: hʌndrəd hed əv kætl

two hundred head of cattle

'two hundred head of cattle' (NB: 'head' is a numeral classifier-like word)

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG}: *No restrictions*

{N,NUM}: *Rare (only nouns in which the singular and plural forms are the same), hence @*

{N,NUM,OBL,CLF}: *Exceptional, hence @*

Number systems

5. Number distinction: *SG/PL (NB: marked on nouns and through subject-verb agreement and determiners)*

6. Obligatoriness of number marking: *Obligatory*

7. Non-numeral quantifiers: *LQ, e.g. meni 'many', as in mæni bɔiz [many boy-PL] 'many boys'*

* Jennifer Sullivan is a native speaker of English, studying Linguistics at the University of Edinburgh, UK. Email: J.N.Sullivan@sms.ed.ac.uk

Classifier systems

8. Noun classes: *Absent (lost)*

9. Numeral classifiers: *Absent (NB: numeral classifier-like word, e.g. two hundred head of cattle)*

Language: Erromangan
Family/Genus: Austronesian/Oceanic
Country/ Macro Area: Vanuatu/Southeast Asia & Oceania
Reference(s): Crowley, T. (1998)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

kilkil haiten wocon viroc

hook one only small

'only one small hook' (1998: 182)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

noki ndehel

coconut three

'three coconuts' (1998: 65)

{N,NUM,NSG}

ovahiven nduru

woman.PL two

'two women' (1998: 66)

4. Restrictions on the patterns in (3) above:

{N,NUM}: *Used generally with no known restrictions (1998: 65)*

{N,NUM,NSG}: *Used only with a small set of nouns (1998: 65-66), hence @.*

Number systems

5. Number distinction: *SG/PL (1998: 61-62)*

6. Obligatoriness of number marking: *Optional (NB: strongly preferred) (1998: 66)*

7. Non-numeral quantifiers: *LQ, e.g. comwisacsu 'all', as in ov-netai comwisacsu [PL-book all] 'all the books' (1998: 168).*

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Absent (WALS)*

Language: Estonian
Family/Genus: Uralic/Finnic
Country/ Macro Area: Estonia/Eurasia
Reference(s): Campbell, G. (2000), Merilin Miljan (personal communication)* and Virve-Anneli Vihman (personal communication)*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

yks *koer*
one *dog[NOM.SG]*
'one dog' (Virve-Anneli Vihman, p.c.)

2. Restrictions on the patterns in (1) above: *No restrictions (Melirin Miljan, p.c.)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,OBL,SG}

viis *koer-a*
five *dog-PRTV.SG*
'five dogs' (Virve-Anneli Vihman, p.c.)

4. Restrictions on the patterns in (3) above: *No restrictions (Melirin Miljan, p.c.)*

Number systems

5. Number distinction: *SG/PL (WALS)*

6. Obligatoriness of number marking: *Obligatory (Melirin Miljan, p.c.)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (2000: 526)*

9. Numeral classifiers: *Absent (Melirin Miljan, p.c.)*

* Merilin Miljan is a native speaker of Estonian, studying Linguistics at the University of Edinburgh, UK. Email: merilin@ling.ed.ac.uk

* Virve-Anneli Vihman (PhD) is an Estonian linguist and Head of International Relations Office at the University of Tartu, Estonia. Email: virve.vihman@ut.ee

Language: Finnish
Family/Genus: Uralic/Finnic
Country/ Macro Area: Finland/Eurasia
Reference(s): Sulkala, K. and M. Karjalainen (1992)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

yksi poika
one boy [NOM.SG]
'one boy' (1992: 345)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,OBL,SG}

kaksi poika-a
two boy-PRTV.SG
'two boys' (1992: 345)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (1992: 345)*

Number systems

5. Number distinction: *SG/PL (WALS)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: *LQ, e.g. monet 'many', as in monet oppilaat [many-PL pupil-PL] 'many pupils' (1992: 96)*

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Absent (WALS)*

Language: French
Family/Genus: Romance/Indo-European
Country/ Macro Area: Romance/Eurasia
Reference(s): Hawkins, R. and Towell, R. (2001)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

un/une *touriste*

one.M/one.F *tourist (M/F)*

'*a/one tourist*' (2001: 8) (*un/une = a, one*, 2001: 126)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions* (2001: 5-9)

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

quatre chats

four cat-PL

'*four cats*' (2001: 126)

{N,NUM,OBL,NSG}

deux cents *millions* *d'habitants*

two hundred-PL *million-PL* *inhabitant.PRTV*

'*two hundred million inhabitants*' (2001: 131)

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG}: *Used generally with no known restrictions* (2001: 126, 131, 138)

{N,NUM,OBL,NSG}: *Used with the numerals million 'million' and milliard 'billion', e.g. cinq cents milliards de francs 'five hundred billion francs' (2001: 131)*

Number systems

5. Number distinction: *SG/PL* (2001: 17)

6. Obligatoriness of number marking: *Obligatory* (2001: 17)

7. Non-numeral quantifiers: *LQ*, e.g. *beaucoup* 'many', *peu* 'few', *tous les* 'all the', e.g. *beaucoup de client-s* [*many of client-PL*] 'many clients' (2001: 143)

Classifier systems

8. Noun classes: *Present* (2001: 5)

9. Numeral classifiers: *Absent* (WALS)

Language: Fur
Family/Genus: Nilo-Saharan/Fur
Country/ Macro Area: Sudan /Africa
Reference(s): Tucker, A. N. and M.A. Bryan (1966)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

murta tok

horse one

'one horse' (1966: 228)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

fuugɔ-ŋa ɔŋal

mountain-PL four

'four mountains' (1966: 228)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1966: 221-222)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (1966: 222)*

9. Numeral classifiers: *Absent (WALS)*

Language: Futuna-Aniwa
Family/Genus: Austronesian/Oceanic
Country/ Macro Area: Vanuatu /Southeast Asia & Oceania
Reference(s): Dryer, M. (1989)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NSG}

ru *fare*

DU *house*

'two houses' (Dougherty 1983 in 1989: 869)

{N,NUM,NSG}

ru *tagata* *e* *rua*

DU *man* *two*

'two men' (Dougherty 1983: 23 in 1989: 869) (NB: *e* = not glossed)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL (1989: 869)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *Absent (WALS)*

Language: Gaagudju
Family/Genus: Australian/Gaagudju
Country/ Macro Area: Australia (Northern Territory)/Australia-New Guinea
Reference(s): Harvey, M. (2002)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

maganaboobu *ngoyoogoda*

banyon *one.F*

'one banyon tree' (2002: 318)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions* (2002: 318)

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

bardangeeya=mba *garrmanoogodawa*

old.woman=AUG *three.M*

'the three old women' (2002: 281) (NB: AUG used as NSG)

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG}: *Used only with human nouns* (2002: 268)

Number systems

5. Number distinction: *SG/PL* (2002: 281)

6. Obligatoriness of number marking: *Obligatory, only human nouns* (WALS)

7. Non-numeral quantifiers: *LQ, e.g. baalgi 'lots'* (2002: 316)

Classifier systems

8. Noun classes: *Present* (2002: 318)

9. Numeral classifiers: *Absent* (WALS)

Language: Gaelic (Scots)
Family/Genus: Indo-European/Celtic
Country/ Macro Area: United Kingdom/Eurasia
Reference(s): Gillies, W. (1993), Lamb, W. (2001), Wilson McLeod, personal communication*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

aon *bhrog*
one *shoe*
'one shoe' (1993: 172)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

dà *chù*
two *dog*
'two dogs'

mile *cù*
thousand *dog*
'thousand dogs' (Wilson McLeod, p.c.)

{N,NUM,NSG}

tri *coin*
three *dog.PL*
'three dogs' (Wilson McLeod, p.c.)

4. Restrictions on the patterns in (3) above:

{N,NUM} : *Used with the numeral 2, and compound numerals with 1 and 2 (e.g. 11, 21); and with high round numerals such as 20, 100 (2001: 38)*

{N,NUM,NSG}: *Used with the numerals 3-19 with some exceptions noted in {N,NUM} (2001: 38)*

Number systems

5. Number distinction: *SG/PL (1993: 172)*

6. Obligatoriness of number marking: *Obligatory (Wilson McLeod, p.c.)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (WALS)*

9. Numeral classifiers: *Absent (1993: passim)*

* Wilson McLeod (PhD) is Lecturer in Celtic and Scottish Studies at the University of Edinburgh, UK. Email: wmcLeod@staffmail.ed.ac.uk

Language: Georgian
Family/Genus: Kartvelian/Kartvelian
Country/ Macro Area: Georgia /Eurasia
Reference(s): Harris, A.C. (1981) and Hewitt, B.G. (1995), (1996)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

ert-i *tve-a*
one-NOM *month-NOM*
'one month' (1995: 668)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions* (1995: 655; 1995:668)

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

sami *knuti*
three *kitten*
'three kittens' (1981:22)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL* (1981: 20)

6. Obligatoriness of number marking: *Obligatory* (WALS)

7. Non-numeral quantifiers: *LQ*, e.g. *mraval* 'several', as in *mraval-ma k'ac-ma* [*several-ERG man.PL-ERG*] 'several men' (1995: 154)

Classifier systems

8. Noun classes: *Absent* (WALS)

9. Numeral classifiers: *Absent* (WALS)

Language: German
Family/Genus: Indo-European/Germanic
Country/ Macro Area: Austria, Germany, Switzerland/Eurasia
Reference(s): Carsten Stuber, personal communication*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

ein *hund*
one *dog*
'one dog'

2. Restrictions on the patterns in (1) above: *No restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

zwei *hunde*
two *dog.PL*
'two dogs'

{N,NUM}

zwei *löffel*
two *spoon*
'two spoons'

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG}: *No restrictions*

{N,NUM}: *Used with some nouns, e.g. those ended in -er or -el, hence @*

Number systems

5. Number distinction: *SG/PL*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: *LQ, e.g. manche 'some', as in manche Frauen [some woman.PL] 'some women'*

Classifier systems

8. Noun classes: *Present (WALS)*

9. Numeral classifiers: *Absent (WALS)*

* Carsten Stuber is a native speaker of German and a visiting student of Physics at the University of Edinburgh, UK. Email: s0347953@sms.ed.ac.uk

Language: Great Andamanese
Family/Genus: Andamanese/Great Andamanese
Country/ Macro Area: India (Andaman Islands)/Southeast Asia & Oceania
Reference(s): Anvita Abbi (personal communication)²; Manoharan, S. (1989)

CNNCs

1. Structural patterns of CNNC_{SG}:

{*N,NUM*}

cokbi *ondoplɔ*

turtle *one*

'one turtle' (1989: 81)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}

{*N,NUM*}

cokbi *oŋjinkɔ*

turtle *two*

'two turtles' (1989: 81)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction* (1989: 61)

6. Obligatoriness of number marking: *See* (5)

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent* (Anvita Abbi, p.c.)

9. Numeral classifiers: *No information*

² Anvita Abbi is Professor of Linguistics at Jawaharlal Nehru University, India. Email: anvitaabbi@hotmail.com

Language: Grebo
Family/Genus: Niger-Congo/Kru
Country/ Macro Area: Liberia /Africa
Reference(s): Innes, G. (1966)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

kae do'

house one

'one house' (1966: 88)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

kia s̄

house.PL two

'two houses' (1966: 88)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1966: 88)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (WALS)*

9. Numeral classifiers: *No information*

Language: Greek (Modern)
Family/Genus: Indo-European/Greek
Country/ Macro Area: Greece /Eurasia
Reference(s): Giorgos Argyropoulos (personal communication)⁸

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

'ena pe'thi
one child
'one child'

2. Restrictions on the patterns in (1) above: *No restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

'thio peth'ia
two child.PL.NOM
'two children'

4. Restrictions on the patterns in (3) above: *No restrictions*

Number systems

5. Number distinction: *SG/PL (WALS)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: *LQ, e.g. po'la 'many', as in po'la pe'dia [many.PL.N.NOM child.N.PL.NOM] 'many children'*

Classifier systems

8. Noun classes: *Present (WALS)*

9. Numeral classifiers: *Absent*

⁸ Giorgos Argyropoulos is a native speaker of Greek, studying Linguistics at the University of Edinburgh, UK. Email: giorgos@ling.ed.ac.uk

Language: Haida
Family/Genus: Haida/Haida
Country/ Macro Area: Canada, United States (Alaska)/North America
Reference(s): Hori, H. (2001)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

qwaay *SGa-SGwansəŋ*

rope *CLF-be.one*

'one piece of rope' (2001: 145)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

qwaay *SGa-sdiŋ- gən*

rope *CLF-be.two-PST*

'There were two pieces of rope.' (2001: 144)

{N,NUM}

qwaay *sdiŋ-gən*

rope *be.two-PST*

'There were two pieces of rope.' (2001: 145)

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF}: *No known restrictions, less frequently observed than {N,NUM}(2001: 144-145), hence @*

{N,NUM}: *No known restrictions, more frequently observed (2001: 144-145)*

Number systems

5. Number distinction: *No information*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Optional (2001: 144-145)*

Language: Halkomelem (Musqueam)
Family/Genus: Salishan/Central Salishan
Country/ Macro Area: Canada /North America
Reference(s): Suttles, W. (2004)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

nə́caʔ-əs *smé·nt*

one-round object *rock*

'one rock/another rock'

{N,NUM}

nə́caʔ ʔəx^wín smə́yəθ

one little deer

'one little deer'

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNCNSG:

{N,NUM,CLF,NSG}

t'xə́mələ *stənté́nəy'*

six.CLF (lit. 'person') *woman.PL*

'six women' (2004: 66)

{N,NUM}

isə́lə *mə́lləs*

two *raccoon*

'two raccoons' (2004: 66)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2004: 204)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Present (2004: 464-465)*

Language: Hatam
Family/Genus: West Papuan/Hatam
Country/ Macro Area: Indonesia (Irian Jaya)/Australia-New Guinea
Reference(s): Reesink, G.P. (1999)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>nab</i>	<i>ni-ndig</i>	<i>can</i>	<i>di-ma</i>
<i>pig</i>	<i>3SG-big</i>	<i>two</i>	<i>REL-that</i>

'those two big pigs' (1999: 58)

{N,NUM,CLF}

<i>nab</i>	<i>ni-ngud</i>	<i>can</i>	<i>ni-ndig</i>	<i>di-ma</i>
<i>pig</i>	<i>3SG-CLF</i>	<i>two</i>	<i>3SG-big</i>	<i>REL-that</i>

'those two big pigs' (1999: 58)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (only animate nouns)(1999: 50)*
6. Obligatoriness of number marking: *Optional (1999: 50)*
7. Non-numeral quantifiers: *LQ, e.g. mang 'many', as in ig mang [house many] 'many houses' (1999: 73)*

Classifier systems

8. Noun classes: *Absent (WALS)*
9. Numeral classifiers: *Optional (WALS)*

Language: Hausa
Family/Genus: Afro-Asiatic/West Chadic
Country/ Macro Area: Niger, Nigeria/Africa
Reference(s): Schuh, R. G. (1999) and Smirnova, M. (1982)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

saniya *d'aya*

cow *one*

'one cow' (1999: 1)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}

{N,NUM,NSG}

shanu *biyu*

cow.PL *two*

'two cows' (1999: 1)

{N,NUM}

gida *goma*

house *ten*

'ten houses' (1999: 1)

{N,NUM,CLF}

doki *guda* *biyar*

horse CLF (lit. 'piece) *five*

'five horses' (1982: 39) (NB: *guda* = numeral classifier-like word)

4. Restrictions on the patterns in (3) above

{N,NUM,NSG}: *Used with animate nouns (1999: 1)*

{N,NUM}: *Used with inanimate nouns (1999: 1)*

{N,NUM,CLF}: *Occasional (1982), hence @*

Number systems

5. Number distinction: *SG/PL (1999: 1)*

6. Obligatoriness of number marking: *Obligatory, only animate nouns (1999: 1)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (WALS)*

9. Numeral classifiers: *Absent (NB: numeral classifier-like words, see (3))*

Language: Hawaiian Creole
Family/Genus: Creoles and Pidgins
Country/ Macro Area: United States (Hawaii)/Southeast Asia & Oceania
Reference(s): Velupillai, V. (2003) and personal communication*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

wan leg

one leg

'one leg' (2003: 174)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (Velupillai 2003: 61, 174)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

tu leg-s

two leg-PL

'two legs' (2003: 174)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (Viveka Velupillai, p.c.)*

Number systems

5. Number distinction: *SG/PL (2003: passim)*

6. Obligatoriness of number marking: *Obligatory (Viveka Velupillai, p.c.)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (Viveka Velupillai, p.c.)*

9. Numeral classifiers: *Absent (Viveka Velupillai, p.c.)*

* Viveka Velupillai (PhD) is Lecturer in Linguistics at the Institut für Anglistik, Giessen, Germany.
Email: viveka.velupillai@email.de

Language: Hebrew
Family/Genus: Afro-Asiatic/Semitic
Country/ Macro Area: Israel/Africa
Reference(s): Ojeda, A. E. (1994)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

yom *exad*

day *one*

'one day' (1994: 1)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}

{N,NUM,NSG}

shlosa *yamim*

three *day.PL*

'three days' (1994: 1)

{N,NUM}

axad-asar *yom*

eleven *day*

'eleven days' (1994: 1)

4. Restrictions on the patterns in (3) above

{N,NUM,NSG}: Used with the numerals 3-10 (1994: 1)

{N,NUM}: Used with the numerals greater than 10 (1994: 1)

Number systems

5. Number distinction: *SG/DU/PL (1994: 1)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (WALS)*

9. Numeral classifiers: *Absent (WALS)*

Language: Hmong Njua
Family/Genus: Hmong-Mien/ Hmong-Mien
Country/ Macro Area: China/Southeast Asia & Oceania
Reference(s): Lyman, T. A. (1979)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

ʔ. lú naŋgâo

one CLF great.boat

'one great boat' (1979: 48)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

pé lú cě

three CLF house

'three houses' (1979: 21)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction (1979: passim)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Present (1979: 21)*

Language: Hungarian
Family/Genus: Uralic/Ugric
Country/ Macro Area: Hungary/Eurasia
Reference(s): Kenesei, I., Vago, R. M. and Fenyvesi, A. (1998) and Victor Tron (personal communication)*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

egy kutya

one dog

'one dog' (Victor Tron, p.c.)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

két kutya

two dog

'two dogs' (Victor Tron, p.c.)

{N,NUM,CLF}

egy szál cigaretta

one CLF (lit. 'piece') cigarette

'one cigarette' (Victor Tron, p.c.)

4. Restrictions on the patterns in (3) above:

{N,NUM}: *Used generally with no known restrictions (1998: 220, 343)*

{N,NUM,CLF}: *Used only with some nouns (Victor Tron, p.c.), hence @*

Number systems

5. Number distinction: *SG/PL (1998: 254)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: *LQ, e.g. sok 'many', as in sok naptár [many calendar] 'lots of calendars' (1998: 230)*

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Optional (WALS)*

* Victor Tron is a native speaker of Hungarian, studying Linguistics at the University of Edinburgh, UK. Email: v.tron@ed.ac.uk

Language: Hunzib
Family/Genus: Nakh-Daghestanian/Avar-Andic-Tsezic
Country/ Macro Area: Russia/Eurasia
Reference(s): Berg, H. (1995)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>be.λ.no</i>	<i>qoqo</i>	<i>li</i>
<i>eight</i>	<i>house (V)</i>	<i>be (V)</i>

'There are eight houses' (1995: 36)

{N,NUM,NSG}

<i>λa.na</i>	<i>q'əra</i>	<i>lo</i>
<i>three</i>	<i>child</i>	<i>be.PL</i>

'Three children' (1995: 36)

4. Restrictions on the patterns in (3) above:

{N,NUM}: *Used generally with no known restrictions (1995: 36)*

{N,NUM,NSG}: *Exceptional (1995: 36), hence @*

Number systems

5. Number distinction: *SG/PL (1995: 35)*
6. Obligatoriness of number marking: *Obligatory (WALS)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (WALS)*
9. Numeral classifiers: *Absent (WALS)*

Language: Iau
Family/Genus: Lakes Plain/Lakes Plain
Country/ Macro Area: Indonesia/Australia-New Guinea
Reference(s): Bateman, J. (1986)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

*ty*⁷ *bi*⁷*si*⁹

person *one*

'one person' (1986: 44)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

*ty*⁷ *bo*⁴

person *two*

'two people' (1986: 18)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No information*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Ijo (Kolokuma)
Family/Genus: Niger-Congo/Ijoid
Country/ Macro Area: Nigeria/Africa
Reference(s): Williamson, K. (1965)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

kpé kẹ̀nì fun

only one book

'only one book' (1965: 92)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

kpékpe ma fùn-kumọ̀

only two book-only

'only two books' (1965: 92)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (WALS)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (Lost) (1965: 6)*

9. Numeral classifiers: *No information*

Language: Ika
Family/Genus: Chibchan/Aruak
Country/ Macro Area: Colombia /South America
Reference(s): Frank, P. (1990)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

in ʔgui *tšeirua*

one *man*

'one man' (1990: 33)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

mouga *tšeirua*

two *man*

'two men' (1990: 32)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction* (1990: 29)

6. Obligatoriness of number marking: *See* (5)

7. Non-numeral quantifiers: *LQ, e.g. imɿ, as in perɿ imɿ [dog many] 'many dogs'* (1990: 29)

Classifier systems

8. Noun classes: *Absent* (WALS)

9. Numeral classifiers: *Absent* (WALS)

Language: Imonda
Family/Genus: Border/Border
Country/ Macro Area: Papua New Guinea /Australia-New Guinea
Reference(s): Seiler, W. (1985)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

mugasl *toad*
one *fellow*
 'one fellow' (1985: 62)

{N,NUM,NPL}

toad-ianèi *mugasl*
boy.PL-NPL *one*
 'one boy' (1985: 62)

2. Restrictions on the patterns in (1) above:

{N,NUM}: No known restrictions

{N,NUM,NPL}: Used with 5 nouns, namely *agō* 'women'; *id* 'men'; *toad* 'boys'; *modòd* 'girls'; *ōgòt* 'enemies' (1985: 62), hence @

3. Structural patterns of CNNC_{NSG}:

{N,NSG}

agō *-nēgle* *-pef*
woman *-DU* *-POSS*
 'your two wives' (1985: 36)

{N,NUM}

mo *-pef* *sabla*
daughter *-POSS* *two*
 'your two daughters' (1985: 37)

{N,NUM,NPL}

agō *-ianèi* *sabla*
woman.PL *-NPL* *two*
 'two women' (1985: 39)

4. Restrictions on the patterns in (3) above:

{N,NSG} and {N,NUM}: Used generally with no known restrictions

{N,NUM,NPL}: Used with 5 nouns, namely *agō* 'women'; *id* 'men'; *toad* 'boys'; *modòd* 'girls'; *ōgòt* 'enemies' (1985: 62), hence @

Number systems

5. Number distinction: SG/DU/PL (NB: kinship terms only) (1985: 36)

6. Obligatoriness of number marking: Optional (1985: 36)

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Absent (WALS)*

Language: Inanwatan
Family/Genus: Marind/South Bird's Head
Country/ Macro Area: Indonesia/Australia-New Guinea
Reference(s): de Vries, L. (2004)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

mésida-e rag-e

man-M one-M

'one man' (2004: 66)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

duqúrewé ériwo

bird.PL two

'two birds' (2004: 61)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2004: passim)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (2004: passim)*

9. Numeral classifiers: *No information*

Language: Iquito
Family/Genus: Zaparoan/Zaparoan
Country/ Macro Area: Peru/South America
Reference(s): Brown, M. C. (2004)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N, NUM}

<i>amicaaca</i>	<i>qui-cumi</i>	<i>iimina</i>	<i>mii</i>	<i>-rĥ</i>
<i>tomorrow</i>	<i>ISG-two</i>	<i>canoe</i>	<i>make</i>	<i>-MPA</i>

'Tomorrow I will make two canoes.' (2004: 63)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2004: 24)*
6. Obligatoriness of number marking: *Optional (2004: 25)*
7. Non-numeral quantifiers: *LQ, e.g. masiana 'many', as in masiana tii iina misi [many COP DET cat] 'There are many cats.'* (2004: 28)

Classifier systems

8. Noun classes: *Absent (2004: 24)*
9. Numeral classifiers: *No information*

Language: Irish
Family/Genus: Indo-European/Celtic
Country/ Macro Area: Ireland /Eurasia
Reference(s): Doyle, A. (2001), Mac Eoin, G. (1993/2002), Ó Dochartaigh, C. (1992) and Ó Siadhail, M. (1989)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

aon mhadadh (amháin)

one dog (only)

'one dog' (1992: 55)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

dhá bhád

two boat

'two boat' (1992: 55)

{N,NUM,NSG}

dhá mhuic

two pig-DU

'two pigs' (1992: 55)

4. Restrictions on the patterns in (3) above:

{N,NUM}: *No restrictions (2001: 55)*

{N,NUM,NSG}: *No known restrictions and used alternatively in modern Irish (1993/2002: 117)*

Number systems

5. Number distinction: *SG/PL (WALS)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: *LQ, e.g. cúpla 'a couple, a few', as in cúpla duine [a.few person] 'a few people' (2001: 98)*

Classifier systems

8. Noun classes: *Present (1989: 143)*

9. Numeral classifiers: *Absent (1989: passim)*

Language: Itzaj
Family/Genus: Mayan/Mayan
Country/Macro Area: Guatemala/North America
Reference(s): Hofling, C. A. (2000)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

jun=ku(u) naj
one=CLF house
'one house' (2000: 24)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

ox=kuul kum
three=CLF pot
'three pots' (2000: 197)

{N,NUM,CLF,NSG}

a' ox=tuul mejen paal-oo'- ej
DET three=CLF small child-PL-TOP
'The three small children' (2000: 228)

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF}: *No known restrictions*

{N,NUM,CLF,NSG}: *Used when "the NP encode specific, given information", e.g. 'my two children' 'my' is specific information. Normally, PL is optional if the NP contains a numeral (2000: 227-228), hence @*

Number systems

- 5. Number distinction: *SG/PL (2000: 227)*
- 6. Obligatoriness of number marking: *Optional (2000: 227-228)*
- 7. Non-numeral quantifiers: -

Classifier systems

- 8. Noun classes: *No information*
- 9. Numeral classifiers: *Present (2000: passim)*

Language: Jamaican Creole
Family/Genus: Creoles and Pidgins
Country/ Macro Area: Jamaica/North America
Reference(s): Bailey, B.L. (1966), Le Page, R.B. and De Camp, D. (1960) and Patrick, P.L. (2003) and personal communication*

1. Structural patterns of CNNC_{SG}:

{N,NUM}

wan son

one son

'one son' (1960: 143)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

di tu book

DEF.DET two book

'the two books' (1966: 30)

{N,NUM,NSG}

di trii sistaz

DEF.DET three sister.PL

'the three sisters' (1960: 143)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL* (2003: 32-34)

6. Obligatoriness of number marking: *Optional* (2003: 33, *passim*)

7. Non-numeral quantifiers: *LQ*, e.g. *tuu* 'a few', as in *tuu brik* [*a.few brick*] 'a few bricks' (2003: 28)

Classifier systems

8. Noun classes: *Absent* (Peter Patrick, *p.c.*)

9. Numeral classifiers: *Absent* (Peter Patrick, *p.c.*)

* Peter Patrick is Professor of Linguistics at the University of Essex. UK. Email: patrickp@essex.ac.uk

Language: Japanese
Family/Genus: Japanese/Japanese
Country/ Macro Area: Japan /Eurasia
Reference(s): Iwasaki, S. (2002) and Kyoko Otsuki (personal communication)*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

<i>hi-tori-no</i>	<i>josee</i>
<i>one-CLF-COP.ATTR</i>	<i>woman</i>
<i>'one woman' (2002: 179)</i>	

2. Restrictions on the patterns in (1) above: *No restrictions (Kyoko Otsuki, p.c.)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

<i>ni-satsu-no</i>	<i>hon</i>
<i>two-CLF-COP.ATTR</i>	<i>book</i>
<i>'two books' (2002: 179)</i>	

4. Restrictions on the patterns in (3) above: *No restrictions (Kyoko Otsuki, p.c.)*

Number systems

5. Number distinction: *No distinction (2002: 53)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: *Reduplication, e.g. yama (SG)/yama-yama (PL) 'mountain' (2002: 78); Person Plural/Group Suffixes, e.g.*

<i>-gata (honorific)</i>	<i>sensee-gata 'teachers'</i>
<i>-tachi</i>	<i>kodomo-tachi 'children'</i>
<i>-ra</i>	<i>aitsu-ru 'those guys' (2002: 72)</i>

Classifier systems

8. Noun classes: *Absent (2002: 2)*

9. Numeral classifiers: *Obligatory (WALS)*

* Kyoko Otsuki is a native speaker of Japanese, studying Linguistics at the University of Edinburgh, UK. Email: K.Otsuki@sms.ed.ac.uk

Language: Jaqaru
Family/Genus: Aymaran/Aymaran
Country/ Macro Area: Peru /South America
Reference(s): Hardman, M. J. (1966) and (2000)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

may *uta*
one *house*
'one house' (2000: 38)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

paj *uta*
two *house*
'two houses' (2000: 38)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (NB: only human nouns) (2000: 11)*

6. Obligatoriness of number marking: *Optional (2000: 11)*

7. Non-numeral quantifiers: *Reduplication "[...] is used for emphasis or for indicating plurality", e.g. quĉ.quĉa 'many lakes' (toponym) (1966: 112)*

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Absent (WALS)*

Language: Jarawara
Family/Genus: Aruan/Aruan
Country/ Macro Area: Brazil/South America
Reference(s): Dixon, R.M.W. (2004)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}

{N,NUM}

jobe_S joro ni-kime
house.M stand AUX-two.M

'two houses stand (there).' (NB: ⁻ⁱkima = suffix meaning two participants, a pair) (2004: 152)⁵

{N,NUM,NSG}

hijama_O mee otaa tao ka-ni-kima-mina otaa-ke
peccary.M 3NSG IEXA shoot APPL-AUX-two-morning.F IEXC-DECF

'We shot two white-lipped peccaries this morning' (2004: 155) ('mee' used as NSG, 2004: 302)

4. Restrictions on the patterns in (3) above:

{N,NUM}: *Used with inanimate nouns* (2004: 152)

{N,NUM,NSG}: *Used with animate nouns* (2004: 155)

Number systems

5. Number distinction: SG/DU/PL (*only animate nouns*) (2004: 75, 261)
6. Obligatoriness of number marking: *Obligatory* (2004: 75)
7. Non-numeral quantifiers: LQ, e.g. *-tama* 'be.many' (2004: *passim*)

Classifier systems

8. Noun classes: *Present* (2004: 152, 279, *passim*)
9. Numeral classifiers: *No information*

⁵ "It is likely that before contact with Branco culture, the Jarawara did not indulge in counting and did not use lexical numbers. A modern-day conversation such as 'How many fish did you catch?' 'Two (or three or seven or eleven)' simply did not occur." (2004: 559)

Language: Jemez
Family/Genus: Kiowa-Tanoan/Kiowa-Tanoan
Country/ Macro Area: United States (New Mexico)/North America
Reference(s): Yumitani, Y. (1998)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM(INV)}

<i>wî</i>	<i>ḿ̥·sá-š</i>	<i>t̥il-kʲy̥ â·</i>
<i>two</i>	<i>cat-INV</i>	<i>TR.1SG.3DU.-put.down.PFV</i>

'I put down two cats.' (1998: 150)

<i>tá</i>	<i>ḿ̥·sá-š</i>	<i>te-gʲó·sê</i>
<i>three</i>	<i>cat-INV</i>	<i>TR.1SG.3INV.-put.down.PFV</i>

'I put down three cats.' (1998: 150)

4. Restrictions on the patterns in (3) above: *Same as (2)*

{N,NUM}: *Used with the nouns where basic number is singular (1998: 97-100)*

{N,NUM,INV}: *Used with the nouns where basic number is dual or plural (1998: 97-100)*

Number systems

5. Number distinction: *SG/DU/PL (1998: 97-100)*
6. Obligatoriness of number marking: *Obligatory (1998: 97)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *Absent (1988: passim)*

Language: Kaki Ae
Family/Genus: Eleman/Eleman
Country/ Macro Area: Papua New Guinea /Australia-New Guinea
Reference(s): Clifton, J.M. (1997)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

ara'a oki

canoe one

'one canoe' (1997: 31)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

ahara-fe ü'üka

brother-PL two

'two brothers' (1997: 25)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1997: 23)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ, e.g. himiri 'many', as in aiparo himiri [pig many] 'many pigs' (1997: 22)*

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Kambera
Family/Genus: Austronesian/Central-Malayo-Polynesian
Country/ Macro Area: Indonesia /Southeast Asia & Oceania
Reference(s): Klamer, M. (1998) and personal communication¹

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

hau *kajawa*
one.CLF *papaya*
 'one papaya' (1998: 139)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (1998: 139)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

tailu *mbua* *kajawa*
three *CLF* *papaya*
 'three papayas' (1998: 93)

{N,NUM,RMS}

tau *ma-dua*
person *RMS-two*
 'two persons/people' (lit. 'people that are two') (1998: 139)

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF}: *Used generally but not with human nouns (1998: 139)*

{N,NUM,RMS}: *Used only with human nouns (1998: 139)*

Number systems

5. Number distinction: *No distinction (Marian Klamer, p.c.)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Obligatory (WALS)*

¹ Marian Klamer (PhD) is Lecturer in Linguistics at the University of Leiden, the Netherlands. Email: M.A.F.Klamer@let.leidenuniv.nl

Language: Kana
Family/Genus: Niger-Congo/Cross River
Country/ Macro Area: Nigeria/Africa
Reference(s): Ikoru, S. M. (1994)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

<i>zii</i>	<i>kā</i>	<i>wā</i>
<i>one</i>	<i>CLF</i>	<i>wife</i>

'one wife' (1994: 17)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (1994: 13, passim)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

<i>bàè</i>	<i>kā</i>	<i>wā</i>
<i>two</i>	<i>CLF</i>	<i>wife</i>

'two wives' (1994: 21)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (1994: 13, passim)*

Number systems

5. Number distinction: *No distinction (1994: 11)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (Lost) (1994: 24)*

9. Numeral classifiers: *Obligatory (1994: passim)*

Language: Kanuri
Family/Genus: Nilo-Saharan/Saharan
Country/ Macro Area: Chad, Niger, Nigeria, Sudan/Africa
Reference(s): Lukas, J. (1937)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

kâm tiló

man one

'one man'

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}: *No information*

4. Restrictions on the patterns in (3) above: *No information*

Number systems

5. Number distinction: *SG/PL (WALS)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Absent (WALS)*

Language: Kariri
Family/Genus: Macro-Ge/Kariri
Country/ Macro Area: Brazil /South America
Reference(s): Rodrigues, A. D. (1999)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

bu-bihe ezumu

CLF-one squash

'one squash' (1990: 192)

{N,NUM}

bihe tupã

one god

'one god' (1990: 192)

2. Restrictions on the patterns in (1) above:

"In the Karirí languages the numerals carry a classifying prefix in agreement with the class of phrase head [...]but take no prefix if the head noun is unclassified". (1990: 192)

3. Structural patterns of CNNC_{NSG}: *No information*

4. Restrictions on the patterns in (3) above: *No information*

Number systems

5. Number distinction: *No information*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Present (1999: 192)*

Language: Karó (Arára)
Family/Genus: Tupian/Ramarama
Country/ Macro Area: Brazil/South America
Reference(s): Gabas, N. (1999)

CNNCs

- 1. Structural patterns of CNNC_{SG}: *No information*
- 2. Restrictions on the patterns in (1) above: *No information*
- 3. Structural patterns of CNNC_{NSG}:

{N,NUM }

<i>ma ʔwɪt</i>	<i>ip</i>	<i>ʔy-t</i>	<i>cagárokōnnen</i>	<i>matet</i>
<i>man</i>	<i>fish</i>	<i>catch-IND</i>	<i>two.ADVZ</i>	<i>yesterday</i>

'The man caught two fish yesterday.' (1999: 173)

- 4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

- 5. Number distinction: *Generally, no distinction (1999: 52), but there exists a plural clitic, e.g. kaʔa=toʔ*

[house=PL]'houses' (1999: 95)

- 6. Obligatoriness of number marking: *See (5)*

- 7. Non-numeral quantifiers: *LQ, e.g. paʔpik 'many', as in agóaʔpət péŋ wī-n paʔpik=tem*
[shaman white.man kill-IND many=ADVZ] 'The shaman killed many white men.' (1999: 171)

Classifier systems

- 8. Noun classes: *Absent (1999: 52)*
- 9. Numeral classifiers: *No information*

Language: Kayardild
Family/Genus: Australian/Tangkic
Country/ Macro Area: Australia (Queensland)/Australia-New Guinea
Reference(s): Evans, N. (1995)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}

{N,NUM}

kiyarrng-ka kularrin-da
two-NOM sister-NOM
'two sisters' (1995: 184)

{N,NSG}

mar-iyarrng
hand-DU
'two front flippers' (1995: 184)

4. Restrictions on the patterns in (3) above
{N,NUM}: *No known restrictions*
{N,NSG}: *Used generally with no known restrictions (1995: 235-236)*

Number systems

5. Number distinction: *SG/DU/PL (1995: 183)*
6. Obligatoriness of number marking: *Optional (1995: 183)*
7. Non-numeral quantifiers: *LQ, e.g. mutha 'many', as in mutha-a dangka-a [many-NOM person-PL] 'many people' (p.236)*

Classifier systems

8. Noun classes: *Absent (WALS)*
9. Numeral classifiers: *Absent (2004: passim)*

Language: Ket
Family/Genus: Yeniseian/ Yeniseian
Country/ Macro Area: Russia (Siberia)/Eurasia
Reference(s): Rijkhoff, J. (2002) and Vajda, E. J. (2004)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

<i>qok</i>	<i>qi'm</i>
<i>one</i>	<i>woman</i>

'one woman' (2004: 36)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>qo'm</i>	<i>qim-n</i>
<i>ten</i>	<i>woman-PL</i>

'ten women' (2004: 36)

4. Restrictions on the patterns in (3) above: *No restrictions* "Apart from a small group of nouns which do not appear in the plural (such as the equivalents of mosquito, fish, duck), Ket nouns normally carry the plural suffix when modified by a numeral" (2002: 36)

Number systems

5. Number distinction: *SG/PL* (2004: 19)

6. Obligatoriness of number marking: *Obligatory* (WALS)

7. Non-numeral quantifiers: *LQ*, e.g. *bɛ́ldè* 'all', as in *bɛ́ldè kəʔt* [all kid.PL] 'all kids' (2004: 26)

Classifier systems

8. Noun classes: *Present* (WALS)

9. Numeral classifiers: *Absent* (2004: *passim*)

Language: Khmer
Family/Genus: Austro-Asiatic/ Khmer
Country/ Macro Area: Cambodia /Southeast Asia & Oceania
Reference(s): Jacob, J. M. (1965), (1990) and Huffman, F. (1967)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

<i>tr ɣy</i>	<i>mùəy</i>	<i>kɔntùỳ</i>
<i>fish</i>	<i>one</i>	<i>CLF (lit. 'tail')</i>

'one fish' (1965:146)

{N,NUM}

<i>chkae</i>	<i>mùəy</i>
<i>dog</i>	<i>one</i>

'one dog' (1965:147)

2. Restrictions on the patterns in (1) above:

{N,NUM,CLF}: Used in careful speech or written language (1965: 145)

{N,NUM}: Used in colloquial speech (1965: 145)

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

<i>mənùs</i>	<i>bu:ən</i>	<i>nèək</i>
<i>mankind</i>	<i>four</i>	<i>CLF</i>

'four people'(1990: 84)

{N,NUM}

<i>chkae</i>	<i>b ɣy</i>
<i>dog</i>	<i>three</i>

'three dogs'(1990: 84)

4. Restrictions on the patterns in (3) above: Same to (2)

Number systems

5. Number distinction: *Absent* (1990: *passim*)

6. Obligatoriness of number marking: *See* (5)

7. Non-numeral quantifiers: *LQ*, e.g. *craən* 'many', as in *bɔɔŋ-pqoon craən n̄eq* [older-and-younger- siblings many person] 'many brothers and sisters.'; Reduplication, e.g. *srəy (SG)/srəy-srəy* 'woman' (1967: 132)

Classifier systems

8. Noun classes: *Absent* (WALS)

9. Numeral classifiers: *Optional* (1965: 145)

Language: Khmu'
Family/Genus: Austro-Asiatic/Palaung-Khmuic
Country/ Macro Area: Laos/Southeast Asia & Oceania
Reference(s): Preamsirat, S. (1987)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

màt m̀ò:j màt
 eye one eye
 'an eye' (1987: 35)

{N,NUM}

k̀ò:n m̀ò:j
 child one
 'one child' (1987: 36)

2. Restrictions on the patterns in (1) above:

{N,NUM,CLF}: No known restrictions; used more frequently (1987: 35)

{N,NUM}: No known restrictions, used less frequently (1987: 35), hence @

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

k̀ò:n p̀à:r k̀òn
 child two CLF
 'two children' (1987: 34)

4. Restrictions on the patterns in (3) above: No known restrictions

Number systems

5. Number distinction: No distinction (WALS)

6. Obligatoriness of number marking: See (5)

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: No information

9. Numeral classifiers: Optional (1987: 35)

Language: Kilivila
Family/Genus: Austronesian/Oceanic
Country/ Macro Area: Papua New Guinea/Southeast Asia & Oceania
Reference(s): Senft, G. (1996)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

<i>te-tala</i>	<i>tau</i>
<i>CLF-one</i>	<i>man</i>

'one man' (1996: 30)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (1996: 18, 30)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

<i>ke-yu</i>	<i>waga</i>
<i>wooden-two</i>	<i>canoe</i>

'two canoes' (1996: 18)

{N,NUM,CLF,NSG}

<i>te-yu</i>	<i>tauwau</i>
<i>CLF-two</i>	<i>man.PL</i>

'two men' (1996:30)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

- 5. Number distinction: *SG/PL (WALS)*
- 6. Obligatoriness of number marking: *No information*
- 7. Non-numeral quantifiers: -

Classifier systems

- 8. Noun classes: *Absent (WALS)*
- 9. Numeral classifiers: *Obligatory (WALS)*

Language: Koasati
Family/Genus: Muskogean/Muskogean
Country/ Macro Area: United States (Alabama)/North America
Reference(s): Kimball, G. D. (1991)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>ná:ni-ha</i>	<i>pokkól</i>	<i>awáh</i>	<i>tóklo-n</i>	<i>hí:ca-li-:s</i>
<i>man-PL</i>	<i>ten</i>	<i>and</i>	<i>two-SW</i>	<i>see-ISS.PST</i>

'I just saw twelve men.' (1991: 358)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PAU/PL* (NB: *human nouns only and normally marked on verbs*) (1991: 446)
6. Obligatoriness of number marking: *Optional (WALS)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*
9. Numeral classifiers: *Absent (1991: passim)*

Language: Kolami
Family/Genus: Dravidian/Central Dravidian
Country/ Macro Area: India /Eurasia
Reference(s): Subrahmanyam, P.S. (1998)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

okkon mas

one.M man

'one man' (1998: 306)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF,NSG}

pāj jen mās-ur

five CLF.M man-PL

'five men' (1998: 306)

{N,NUM,NSG}

ayd mās-ur

five man-PL

'five men' (1998: 306)

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG}: *Used with native numerals (1998: 306)*

{N,NUM,CLF,NSG}: *Used with numerals borrowed from Marathi (1998: 306)*

Number systems

5. Number distinction: *SG/PL (1998: 306-307)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (1998: 306)*

9. Numeral classifiers: *Optional (1998: 306)*

Language: Kombai
Family/Genus: Trans-New Guinea/Awju-Dumut
Country/ Macro Area: Indonesia/Australia-New Guinea
Reference(s): de Vries, L. (1993) and Nichols, J. (1992)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

kho mofenadi

man one

'one man' (1993: 40)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

kho molumo

man two

'two men' (1993: 40)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction (WALS)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: *LQ, biduma 'many', as in makhü muyiyano biduma [dog big.PL many] 'many big dogs' (1993: 39)*

Classifier systems

8. Noun classes: *Absent (1992: 296-297)*

9. Numeral classifiers: *Absent (WALS)*

Language: Korean
Family/Genus: Korean/Korean
Country/ Macro Area: Korea /Eurasia
Reference(s): Chonghyuck Kim (personal communication)², Hae-Sung Jeon (personal communication) and Sohn, H. M. (1994), (1999)³

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

ke han mari
dog one CLF
 'one dog' (Chonghyuck Kim, p.c.)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

ke du mari
dog two CLF
 'two dogs' (Chonghyuck Kim, p.c.)

{N,NUM,CLF,NSG}

haksayng-tul twu-myengi
student-PL two-CLF
 'two students' (Chonghyuck Kim, p.c.)

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF}: *Used generally with no known restrictions (1999: 325)*

{N,NUM,CLF,NSG}: *Used only with human nouns NB: PL optional even with human nouns (Chonghyuck Kim, p.c.), hence @*

Number systems

5. Number distinction: *SG/PL (1994: 218)* (NB: *without plural markers, the noun is transnumeral, e.g. salam may denote 'person' or 'persons'.*) (1994: 269)

6. Obligatoriness of number marking: *Optional (1994: 218)*

7. Non-numeral quantifiers: *LQ, e.g. yele 'many, several', as in chayk yele kwen [book several CLF] 'several books' (1999: 353).*

² Chonghyuck Kim is a native speaker of Korean, studying Linguistics at the University of Delaware, USA. Email: cheesue@UDel.edu

³ Hae-Sung Jeon is a native speaker of Korean, studying Linguistics at the University of Cambridge, UK. Email: hsj24@cam.ac.uk

Classifier systems

8. Noun classes: *Absent* (*Hae-Sung Jeon, p.c.*)

9. Numeral classifiers: *Obligatory* (*WALS*)

Language: Korku
Family/Genus: Austro-Asiatic/Munda
Country/ Macro Area: India/Eurasia
Reference(s): Nagaraja, K. S. (1999)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

mya-ga:w

one-village

'one village' (1999: 91)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

ba:ri-kitab

two-book

'two books' (1999: 34)

{N,NUM,NSG}

aphay-kon-ku

three-child-PL

'three children' (1999: 91, 318)

{N,NSG}

siTa-kiñj

dog-DU

'two dogs'

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL (NB: only animate nouns) (1999: passim)*

6. Obligatoriness of number marking: *Obligatory, only human nouns (WALS)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Absent (1999: 91)*

Language: Koromfe
Family/Genus: Niger-Congo/Gur
Country/ Macro Area: Burkina Faso, Mali/Africa
Reference(s): Rennison, J. R. (1997)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,SG}

kē-ō *dufe*

woman-SG *one*

'one woman' (1997: 306)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

zenA *n āā*

year.PL *four*

'four years' (1997: 305)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (1997: 242, 247)*

Number systems

5. Number distinction: *SG/PL (1997: 198)*

6. Obligatoriness of number marking: *Obligatory (1997: 198)*

7. Non-numeral quantifiers: *LQ, e.g. duru 'all', as in a benna duru [ART man.PL all] 'all men' (1997: 83)*

Classifier systems

8. Noun classes: *Present (1997: 206)*

9. Numeral classifiers: *Absent (1997: 233)*

Language: Koyraboro Senni
Family/Genus: Nilo-Saharan/Songhay
Country/ Macro Area: Mali, Niger/Africa
Reference(s): Heath, J. (1999)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

woy foo

woman one

'a woman; one woman' (1999: 10)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

woy hiŋka

woman two

'two women' (1999: 121)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1999: 121)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *No information*

Language: Kpelle
Family/Genus: Niger-Congo/Western Mande
Country/ Macro Area: Liberia /Africa
Reference(s): Campbell, G. L. (2000) and Welmers, W. E. (1973)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

pére *tɔ̀ nɔ̀*

house *one*

'one house' (1973: 294)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

pére *feerɛ*

house *two*

'two houses' (1973: 294)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2000: 914)*

6. Obligatoriness of number marking: *Optional (2000: 914)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (2000: 914)*

9. Numeral classifiers: *Absent (WALS)*

Language:	Kuna
Family/Genus:	Chibchan/Kuna
Country/ Macro Area:	Colombia, Panama /South America
Reference(s):	Holmer, N. M. (1946)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

tule war-kwena

man CLF-one

'one man' (1946: 190)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

ome war-pockwa

woman CLF-two

'two women' (1946: 190)

{NUM,CLF}

kwa-po

CLF (for nuts)-two

'two nuts' (NB: kwa kwa=nut, nuts) (1946: 190)

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF}: No known restrictions

{NUM,CLF}: seems to be non-dominant "Sometimes no noun is used..." (1946: 190), hence @

Number systems

5. Number distinction: *No distinction (1947: passim)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Present (1947: passim)*

Language: Kuot
Family/Genus: Kuot/Kuot
Country/ Macro Area: Papua New Guinea /Australia-New Guinea
Reference(s): Chung, C-H. and Chung, K-J. (1996) and Linström, E. (2002)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

nomurit magabun

one woman

'one woman' (1996:20)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

naien ma kamilip

three PL yam.PL

'three yams' (1996: 19)

{N,NSG}

Dagar-βien

egg-DU

'two eggs' (1996: 42)

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG}: *No known restrictions*

{N,NSG} (NSG=DU): *Used with the nouns representing two referents (1996:42; 2002: 137)*

Number systems

5. Number distinction: *SG/DU/PL (2002: 130)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: *LQ, e.g. papaluaip 'many', as in papaluaip ma magaulap [many PL woman.PL] 'many women' (1996: 19)*

Classifier systems

8. Noun classes: *Present (2002: 130)*

9. Numeral classifiers: *No information*

Language: Kutenai
Family/Genus: Kutenai/Kutenai
Country/ Macro Area: Canada, United States/North America
Reference(s): Dryer, M. (2005b)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{*N, NUM*}

N= *ʔas-iʔ* *ʔupiʔ-iʔ-ni* *ʔa-qʔumaknik*

IND=*two-PREV* *kill-PASS-IND* *Indian*

'*Two Indians were killed.*' (2005b: 363)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (NB: only human nouns) (WALS)*
6. Obligatoriness of number marking: *Obligatory (WALS)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*
9. Numeral classifiers: *Absent (WALS)*

Language: Kwazá
Family/Genus: Kwaza/Kwaza
Country/ Macro Area: Brazil/South America
Reference(s): Voort, H. (2004)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N, NUM, CLF}

<i>a'xy</i>	<i>aky-'xy</i>
<i>house</i>	<i>two-CLF</i>
<i>'two houses' (2004: 131)</i>	

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction (2004: 213)*
6. Obligatoriness of number marking: *See (5)*
7. Non-numeral quantifiers: *words with numerical interpretation, e.g.*

<i>tei-</i>	<i>'to be one', 'alone'</i>
<i>aky-</i>	<i>'to be two', 'company'</i>
<i>e'mā</i>	<i>'one more/ again', 'without companion'</i>
<i>ele'le</i>	<i>'several/many/very/emphatic (2004: 214)</i>

Classifier systems

8. Noun classes: *Absent (2004: 24)*
9. Numeral classifiers: *Obligatory (WALS)*

Language: Kwerba
Family/Genus: Kwerba/Kwerba
Country/ Macro Area: Indonesia /Australia-New Guinea
Reference(s): de Vries, J. A. and de Vries, S. A. (1997)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

cibicabo abarias

pig one

'one pig' (1997: 28)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

cibicabo nenumwano

pig two

'two pigs' (1997: 29)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL (1997: passim)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Lakhota
Family/Genus: Siouan/Siouan
Country/ Macro Area: United States (North Dakota)/North America
Reference(s): Williamson, J. (1984)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

<i>hoksila</i>	<i>wąži</i>	<i>ki</i>
<i>boy</i>	<i>one</i>	<i>DEF.ART</i>

'the one boy' (1984: 53)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>hoksila</i>	<i>nųp</i>	<i>ki</i>
<i>boy</i>	<i>two</i>	<i>DEF.ART</i>

'the two boys' (1984: 53)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

- 5. Number distinction: *SG/PL (NB: not overtly marked on nouns, but through demonstratives and articles) (1984: 41)*
- 6. Obligatoriness of number marking: *No information*
- 7. Non-numeral quantifiers: -

Classifier systems

- 8. Noun classes: *Absent (1984: 40)*
- 9. Numeral classifiers: *Absent (WALS)*

Language: Lango
Family/Genus: Nilo-Saharan/Nilotic
Country/ Macro Area: Uganda /Africa
Reference(s): Noonan, M. (1992)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

gúlú́ àcél

pot one

'one pot' (1992: 167)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

gúlú́ àryǝ

pot two

'two pots' (1992: 167)

{N,NUM,NSG}

Gwóggî à dǝŋǝ àrǝní

Dog.PL ATTR.PCL big two-this

'these two big dogs' (1992: 156)

4. Restrictions on the patterns in (3) above:

{N,NUM}: *No known restrictions*

{N,NUM,NSG}: *Used with animate nouns (1992: 156, 167)*

Number systems

5. Number distinction: *SG/PL (NB: only animate nouns) (1992: 83)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: *LQ, e.g. pòl 'many', as in gúlú́ à pòl [pot ATTR.PCL many] 'many pots' (1992: 167)*

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Absent (WALS)*

Language: Latvian
Family/Genus: Indo-European/Baltic
Country/ Macro Area: Latvia/Eurasia
Reference(s): Nau, N. (1998)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>tā</i>	<i>es</i>	<i>tur</i>	<i>mācīj-os</i>	<i>četr-us</i>	<i>gad-us,</i>
<i>so</i>	<i>1SG.NOM</i>	<i>there</i>	<i>study.PRT-1SG</i>	<i>four-ACC.M</i>	<i>year-ACC.PL</i>

'So I studied there for four years...' (1998: 15)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1998: passim)*
6. Obligatoriness of number marking: *Obligatory (WALS)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (WALS)*
9. Numeral classifiers: *Absent (WALS)*

Language: Lavukaleve
Family/Genus: Solomons East Papuan/ Solomons East Papuan
Country/ Macro Area: Solomon Islands/Australia-New Guinea
Reference(s): Terrill, A. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>molev</i>	<i>sie</i>
<i>canoe.PL</i>	<i>five</i>
<i>'five canoes' (2003: 89)</i>	

{N,NUM}

<i>Nei</i>	<i>kanal</i>	<i>o</i>	<i>kanamil</i>	<i>enga</i>
<i>coconut</i>	<i>tens.DU</i>	<i>or</i>	<i>tens.PL</i>	<i>three</i>
<i>'twenty or thirty coconuts' (2003: 89)</i>				

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG}: *Used generally with no known restrictions (2003: 89, 141, 478)*

{N,NUM}: *Used with nouns which are not marked for plural (2003: 89)*

Number systems

5. Number distinction: *SG/DU/PL (2003: 106)*
6. Obligatoriness of number marking: *Obligatory (2003: 89)*
7. Non-numeral quantifiers: *LQ, e.g. suni 'all', mail 'a bit' (2003: 67)*

Classifier systems

8. Noun classes: *Present (2003: 141)*
9. Numeral classifiers: *Absent (2003: passim)*

Language: Lele
Family/Genus: Afro-Asiatic/East Chadic
Country/ Macro Area: Chad/Africa
Reference(s): Frajzyngier, Z. (2001)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

ìlú pínà

stick one

'one stick' (2001: 94)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}

{N,NUM,NSG}

kàjé sò

knife.PL two

'two knives' (2001: 95)

{N,NUM}

tídì subù

bird three

'three birds.' (2001: 60)

4. Restrictions on the patterns in (3) above

{N,NUM,NSG}: *Used with nouns denoting large animals, kinship terms, and a few inanimate objects*
(2003: 60)

{N,NUM}: *Used with wild animals, birds, and insects* (2003: 60)

Number systems

5. Number distinction: *SG/PL* (2003: 95, *passim*) (*only human nouns, WALS*)

6. Obligatoriness of number marking: *Obligatory* (WALS)

7. Non-numeral *quantifiers*: *LQ*, e.g. *néy* 'many', as in *bán-wé néy* [man-PL many] 'many men'
(2003: 87)

Classifier systems

8. Noun classes: *Present* (WALS)

9. Numeral classifiers: *Absent* (2003: 60)

Language:	Lepcha
Family/Genus:	Sino-Tibetan/Lepcha
Country/ Macro Area:	Bhutan, India, Nepal/Southeast Asia & Oceania
Reference(s):	Plaisier, H. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

muró-kát

man-one

'one/a man' (2003: 708)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (2003: 708, 710)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

àkup nyet

child two

'two children' (2003: 714)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (2003: 710, 714)*

Number systems

5. Number distinction: *SG/PL (2003: 708)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (2003: 707)*

9. Numeral classifiers: *No information*

Language: Leti
Family/Genus: Austronesian/Central Malayo-Polynesian
Country/ Macro Area: Indonesia/Southeast Asia & Oceania
Reference(s): van Engelenhoven, A. (2004)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM,(NSG)}

püata=pura *vòata=e (ra)*

woman=wild *four=DEX (PL)*

'the four nymphomaniacs' (2004: 184)

4. Restrictions on the patterns in (3) above:

{N,NUM}: *No known restrictions*

{N,NUM,NSG}: *Used with human nouns*

Number systems

5. Number distinction: *SG/PL (2004: 184)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: *Reduplication, e.g. rumlalavnâ rum=la-lavna [REP house=REDUP-big]*
'big houses' (2004: 111)

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *Absent (WALS)*

Language: Lezgian
Family/Genus: Nakh-Daghestanian/Lezgian
Country/ Macro Area: Russia (Azerbaijan) /Eurasia
Reference(s): Haspelmath, M. (1993)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

sa ttar

one tree

'one tree.' (1993: 140, 245)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions* (1993: 143, 245)

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

q'we ttar

two tree

'two tree'(1993: 140, 245)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions* (1993: 142, 245)

Number systems

5. Number distinction: *SG/PL* (1993: 4)

6. Obligatoriness of number marking: *Obligatory* (WALS)

7. Non-numeral quantifiers: *LQ*, e.g. *aq' wan* 'many', as in *aq' wan jaš-ar* [*many age-PL*] 'many years' (1993: 199)

Classifier systems

8. Noun classes: *Absent* (1993: 563)

9. Numeral classifiers: *Absent* (1993: 564)

Language: Limilngan
Family/Genus: Australian/Limilngan
Country/ Macro Area: Australia (Northern Territory)/Australia-New Guinea
Reference(s): Harvey, M. (2001)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

<i>minbulungbulung</i>	<i>ajunini</i>
<i>bird</i>	<i>one</i>

'one minbulungbulung' (2001: 113) (*minbulungbulung* = specific name)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}

{N,NUM,NSG}

<i>aykgurr</i>	<i>arnikgan=arnikgan</i>
<i>two</i>	<i>old woman=old woman</i>

'two old women' (2001: 111)

{N,NUM}

<i>dilimin langan</i>	<i>aykgurr</i>
<i>palm</i>	<i>two</i>

'two palm trees' (2001: 93)

4. Restrictions on the patterns in (3) above

{N,NUM,NSG}: *Used with human nouns (2001: 111)*

{N,NUM}: *No known restrictions*

Number systems

5. Number distinction: *No distinction (NB: human nouns may be optionally marked for plural) (2001: 111)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: *Reduplication, e.g. arnikgan (SG)/arnikgan=arnikgan (PL) 'old woman' (2001: 111)*

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Absent (2001: passim)*

Language: Lithuanian
Family/Genus: Indo-European/Baltic
Country/ Macro Area: Lithuania/Eurasia
Reference(s): Ambrasaz, V (1997)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

vienas *berniukas*

one *boy*

'one boy' (1997: 167, 491)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

keturi *vaikai*

four *child.NOM.PL*

'four children' (1997: 587)

{N,NUM,OBL,NSG}

dėšimt *dienų*

ten *day.GEN.PL*

'ten days' (1997: 587)

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG}: *Used with the numerals 2-9 (1997: 167)*

{N,NUM,OBL,NSG}: *Used with the numerals 10-19 and high round numerals (e.g. 20, 100) (1997: 167)*

Number systems

5. Number distinction: *SG/PL (1997: 101)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (1997: 96)*

9. Numeral classifiers: *Absent (1997: passim)*

Language: Lunda
Family/Genus: Niger-Congo/Bantoid
Country/ Macro Area: Democratic Republic of Congo /Africa
Reference(s): Kawasha, B. K. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,SG}

mu-ntu *wu-mú*

I-person *I-one*

'one person' (2003: 124) (Class mu-/wu- = singular number)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

a-ntu *a-yedí*

II-person *II-two*

'two people' (2003: 124) (Class a- = plural number)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2003: passim)*

6. Obligatoriness of number marking: *Obligatory (2003: 73)*

7. Non-numeral quantifiers: *LQ, e.g. -vúlu, as in ma-tála a-a- ma-vúlu [CL-house CL-POSS CL-many] 'many houses' (2003: 121)*

Classifier systems

8. Noun classes: *Present (2003: 73)*

9. Numeral classifiers: *No information*

Language: Maale
Family/Genus: Afro-Asiatic/Omotic
Country/ Macro Area: Ethiopia/Africa
Reference(s): Amha, A. (2001)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

<i>pétte</i>	<i>máár-ka</i>
<i>one</i>	<i>house-LOC</i>

'in one house' (2001: 285)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>lam ?ó</i>	<i>wáár-á</i>
<i>two</i>	<i>goat.PL-NOM</i>

'two goats.' (2001: 52)

{N,NUM}

<i>lam ?ó</i>	<i>waarí</i>
<i>two</i>	<i>goat-NOM</i>

'two goats.' (2001: 52)

4. Restrictions on the patterns in (3) above *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2001: 52)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Absent (2001: 52)*

Language: Ma'di
Family/Genus: Nilo-Saharan/Moru- Ma'di
Country/ Macro Area: Sudan,Uganda/Africa
Reference(s): Fabb, N. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

tà àl̄

thing one

'one thing.' (2003: 575)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}

{N,NUM}

à ʔú ʔk̄ èrĩ

chicken red two

'two red hens.' (2003: 357)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (NB: only human nouns and occupations) (2003: 98, passim)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Madurese
Family/Genus: Austronesian/Sundic
Country/ Macro Area: Indonesia (Java) /Southeast Asia & Oceania
Reference(s): Davies, W. (1999)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:
{*N,NUM*}
lemaq oreng
five person
'five people' (1999: 13)
4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No information*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: *LQ*, e.g. *kabbhi* 'all', as in *kabbhi buku* [*all book*] 'all the books' (1999: 13); *Reduplication oreng* (SG)/*rengoreng* (PL) 'person' (1999: 13)

Classifier systems

8. Noun classes: *Absent* (1999: 13)
9. Numeral classifiers: *No information*

Language: Maltese
Family/Genus: Afro-Asiatic/Semitic
Country/ Macro Area: Malta/Africa
Reference(s): Borg, A and M.A. Alexander (1997)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

bajda *wafida*

egg *one*

'one egg' (1997: 268)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

tliet *bajdiet*

three *egg.PL*

'three eggs' (1997: 268)

{N,NUM}

taxil *bajda*

twelve *egg*

'twelve eggs' (1997: 268)

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG}: *Used with the numerals 2-10 and compound numerals ending with these digits, e.g. mija u tliet bajdiet [hundred and three egg.PL] 'one hundred and three eggs' (1997: 268)*

{N,NUM}: *Used with the numerals greater than 10 (1997: 268)*

Number systems

5. Number distinction: *SG/DU/PL (1997: 174)*

6. Obligatoriness of number marking: *Obligatory (1997: 174)*

7. Non-numeral quantifiers: *LQ, e.g. fitit 'a little', as in fitit ħbieb [a.little friend.PL] 'a few friends' (1997: 72)*

Classifier systems

8. Noun classes: *Present (1997: 188)*

9. Numeral classifiers: *Absent (1997: 192)*

Language: Malto
Family/Genus: Dravidian/Northern Dravidian
Country/ Macro Area: India /Eurasia
Reference(s): Steever, S. B. (1998)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

maq-ond ōydu

CLF-one cow

'one cow' (1998: 372)

2. Restrictions on the patterns in (1) above:

{N,NUM,CLF}: No known restrictions

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF,NSG}

tīni jen male-r

three CLF man-PL

'three men' (1998: 372)

{N,NUM,CLF}

tīni maq ōydu

three CLF cow

'three cows' (1998: 372)

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF,NSG}: Used with human nouns (1998: 372-373)

{N,NUM,CLF}: Used generally but not with human nouns (1998: 372-373)

Number systems

5. Number distinction: *SG/PL (NB: only human nouns) (1998: 362)*

6. Obligatoriness of number marking: *Obligatory (1998: 362)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (1998: 361)*

9. Numeral classifiers: *Obligatory (1998: 372)*

Language: Mam
Family/Genus: Mayan/Mayan
Country/ Macro Area: Guatemala/North America
Reference(s): England, C. (1983)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

<i>jun</i>	<i>tx'yaan</i>	<i>q'aq</i>
<i>one</i>	<i>dog</i>	<i>black</i>

'one black dog' (1983: 147)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>ajaj</i>	<i>oox</i>	<i>tx'yaan</i>	<i>saq</i>
<i>these</i>	<i>three</i>	<i>dog</i>	<i>white</i>

'these three white dogs.' (1983: 149)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL, normally marked on verb (1983: 144-154)*

6. Obligatoriness of number marking: *Obligatory (1983: passim)*

7. Non-numeral quantifiers: *LQ, e.g. kab' 'some', e.g. kab' xiinaq [two man] 'two/some men' (1983: 147)*

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Manchu
Family/Genus: Altaic/Tungusic
Country/ Macro Area: China (Manchuria)/Eurasia
Reference(s): Campbell, G.L. (2000), Gorelova, L. M. (2002), Mark C. Elliott (personal communication)*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

emu *jui*
one *child*
'one child' (Mark C. Elliott, p.c.)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM.NSG}

juwe *juse*
two *child.PL*
'two children' (Mark C. Elliot, p.c.)

{N,NUM}

juwe *indahuun*
two *dog*
'two dogs' (Mark C. Elliot, p.c.)

{N,NUM,CLF}

ilan *fesin* *loho*
three *CLF* *sword*
'three swords' (2002: 206)

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG} and {N,NUM,CLF}: *No known restrictions*

{N,NUM}: *Used with non-human nouns* (Mark C. Elliot, p.c.)

Number systems

5. Number distinction: *SG/PL* (NB: *only human nouns*) (2002: 134)

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ*, e.g. *ememu* 'many', as in *ememu juse* [*many child-PL*] 'many children' (Mark C. Elliot, p.c.)

* Mark C. Elliot is Professor of Chinese and Inner Asian History at Harvard University, USA. Email: mark_elliott@harvard.edu

Classifier systems

8. Noun classes: *Absent* (2000: 1061)

9. Numeral classifiers: *Present* (2002: 206-207)

Language: Mandarin
Family/Genus: Sino-Tibetan/Chinese
Country/Macro Area: China/Southeast Asia & Oceania
Reference(s): Li, C.N. and Thompson, S.A. (1981) and Lin, H. (2001)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

yí	gè	jīdàn
one	CLF	egg

'one egg' (2001: 107)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (2001: 107; 1981: 96)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

sān	wèi	lǎoshī
three	CLF	teacher

'three teachers' (2001: 107)

{NUM,CLF}

sān	tiān
three	day

'three days' (1981: 105)

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF}: *Used generally with no known restrictions, except nouns denoting measures, e.g. 'day' (2001: 107; 1981: 104, 105)*

{NUM,CLF}: *Used with nouns denoting measures, e.g. 'day' (1981: 105)*

Number systems

5. Number distinction: *No distinction (1981: 11-12)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: *LQ, e.g. jǐ 'few', as in zhèi jǐ mén pào [this few CLF cannon] 'these few cannons' (1981: 105)*

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Obligatory (WALS)*

Language: Mangghuer
Family/Genus: Altaic/Mongolic
Country/ Macro Area: China /Eurasia
Reference(s): Slater, K. W. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

yi-ge chuna

one-CLF wolf

'one wolf' (2003: 343)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

mer liang-tiao

road two-CLF

'two roads.' (2003: 55)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2003: 99, 102)*

6. Obligatoriness of number marking: *Optional (2003: 98)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Obligatory (2003: passim)*

Language: Maori
Family/Genus: Austronesian/Oceanic
Country/ Macro Area: New Zealand /Southeast Asia & Oceania
Reference(s): Bauer, W. (1993)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

<i>te</i>	<i>maaripi</i>	<i>kotahi</i>
ART	knife	one

'one knife' (1993: 496)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions* (1993: 152, 496)

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NM}

<i>eenei</i>	<i>waka</i>	<i>e</i>	<i>rua</i>
these	canoe	NUMPCL	two

'these two canoes' (1993: 262)

{N,NUM,PNM}

<i>toko-rima</i>	<i>oona</i>	<i>tuaakana</i>
PNM-five	GEN.PL.3SG	brother

'He had five older siblings' (1993: 496)

4. Restrictions on the patterns in (3) above:

{N,NUM,NUMPCL}: *Used generally with numerals 2-9 and all forms beginning with these numerals (e.g. rua tekau [two ten] 'twenty')* (NB: numerals 10-19 take no prefix) (1993: 496)

{N,NUM,PNM}: *Used with human nouns (NB: traditional use and increasingly ignored)* (1993: 496), hence @

Number systems

5. Number distinction: *SG/PL (NB: only human nouns, e.g. tangata(SG)/taangata(PL) 'man')* (1993: 352)

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ, e.g. eetahi 'some', as in eetahi pukapuka [some book.REDUP] 'some books'* (1993: 364)

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Absent (1993: passim)*

Language: Mapudungun
Family/Genus: Araucanian/Araucanian
Country/ Macro Area: Chile/South America
Reference(s): Smeets, I. (1950), Zúñiga, F. (2000)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

kiñe *nge*

one *eye*

'one eye' (1950: 192)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (1950: 192; 2000: 15)*

3. Structural patterns of CNNC_{NSG}

{N,NUM}

epu *yall*

two *child.of.man*

'two children' (1950: 192)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (1950: 192; 2000: 34)*

Number systems

5. Number distinction: *SG/DU/PL (NB: shown on verbs) (1950: 223)*

6. Obligatoriness of number marking: *Optional (1950: 224)*

7. Non-numeral quantifiers: *LQ, e.g. alün 'many', as in alün achawall [many hen] 'many hens' (2000: 21)*

Classifier systems

8. Noun classes: *Absent (2000: 16)*

9. Numeral classifiers: *Absent (WALS)*

Language: Matsés
Family/Genus: Panoan/Panoan
Country/ Macro Area: Brazil,Peru/South America
Reference(s): Fleck, D. W. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

cuididi *daëd*

brat *two*

'Two kids' (2003: 558)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2003: 273)*
6. Obligatoriness of number marking: *Optional (p.273)*
7. Non-numeral quantifiers: *LQ, e.g. dadpen 'many', as in chido dadpen [woman many] 'many women.' (2003: 240)*

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *No information*

Language: Maung
Family/Genus: Australian/Iwaidjan
Country/ Macro Area: Australia (Northern Territory)/Australia-New Guinea
Reference(s): Capell, A. (1970)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}

{N,NUM}

nargarg *arargbi*

two *man*

'two men' (1970: 126-128)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (NB: only human nouns) (1970: 52)*
6. Obligatoriness of number marking: *Optional (1970: 52)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (1970: 50)*
9. Numeral classifiers: *Absent (WALS)*

Language: Meithei
Family/Genus: Sino-Tibetan/Kuki-Chin-Naga
Country/ Macro Area: India/Southeast Asia & Oceania
Reference(s): Yashawanta, S. (2000)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:
{*N,NUM*}
əŋaŋ əma
kid four
'four kids' (2000: 78)
4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2000: passim)*
6. Obligatoriness of number marking: *Optional (2000: passim)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*
9. Numeral classifiers: *No information*

Language: Meyah
Family/Genus: East Bird's Head/East Bird's Head
Country/ Macro Area: Indonesia/Australia-New Guinea
Reference(s): Gravelle, G. (2002)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

<i>mongkukar</i>	<i>efaga</i>	<i>or-ins</i>
<i>chicken</i>	<i>CLF</i>	<i>CLF-one</i>

'one chicken' (2002: 130)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

<i>manana</i>	<i>efebi</i>	<i>er-eka</i>
<i>papaya</i>	<i>CLF</i>	<i>CLF-two</i>

'two papayas' (2002: 130)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (NB: only nouns denoting humans, dogs, pigs) (2002: 125)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ, e.g. ofokou 'many', as in mod ofokou [house many] 'many houses' (2002: 125)*

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Obligatory (2002: 129)*

Language: Miskito
Family/Genus: Misumalpan/Misumalpan
Country/ Macro Area: Nicaragua/South America
Reference(s): Campbell, G. L. (2000), Casper, B. and Schlaefer, S. (1944) and
 Conzemius, E. (1929)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>waikna</i>	<i>wal</i>
<i>man</i>	<i>two</i>
<i>'two men'</i>	<i>(1929: 75)</i>

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction (1994: passim)*
6. Obligatoriness of number marking: *See (5)*
7. Non-numeral quantifiers: *Plural word, e.g. nani, as in upla nani [person PL] 'persons' (1944: 2)*

Classifier systems

8. Noun classes: *Absent (2000: 1134)*
9. Numeral classifiers: *Absent (WALS)*

Language:	Mixtec (Alcatlatzala)
Family/Genus:	Oto-Manguan/Mixtecan
Country/ Macro Area:	Mexico/North America
Reference(s):	Zylstra, C. F. (1991)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

ñū *lēṣōn*

one *rabbit*

'one rabbit' (1991: 68)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

ovi *taā*

two *man*

'two men' (1991: 68)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No information*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ, e.g. kwaha, as in kwaha kīti [many animal] 'many animals' (1991: 68)*

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Mochica
Family/Genus: Chimúan/Chimúan
Country/ Macro Area: Peru /South America
Reference(s): Adelaar, W. and Muysken P. C. (2004)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

sop mov
three corvina

'three corvinas' (2004: 349)(NB: corvina = a type of fish popular in Peru)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2004: 348)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *Absent (but NB: "ten units of a particular subclass of nouns. For instance, pong was used for (tens of) people, animals and reeds, and ssop for (tens of) coins or time units [...]") (p.342)*

pak-chokij mang
two-CLF.ten maize

'twenty ears of maize' (2004: 343)

Language: Mocoví
Family/Genus: Guaicuruan/Guaicuruan
Country/ Macro Area: Argentina /South America
Reference(s): Grondona, V. M. (1998/2004)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

doh yale

two man

'two men' (1998: 92) (NB: "[Mocovi]has borrowed all its numerals from Spanish") (1998: 91)

{N,NUM,NSG}

doh yale-r

two man-PL

'two men' (1998: 92)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1998: 92)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *No information*

Language: Mohawk (Akwasasne)
Family/Genus: Iroquoian/Northern Iroquoian
Country/ Macro Area: Canada, United States/North America
Reference(s): Bonvillain, N. (1973), Marianne Mithun (personal communication)²

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

v́ska oháhasela?
one lamp/light
'one lamp,light' (Marianne Mithun, p.c.)

{N,SG}

ska- háhselah
SG- lamp/light
'one lamp,light' (1973: 235)

2. Restrictions on the patterns in (1) above:

{N,NUM}: *Used as marked construction in the language (Marianne Mithun, p.c.), hence @*

{N,SG}: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

tékeni oháhasela?
two lamp/light
'two lamps/ lights' (1973: 234-5)

{N,NSG}

teka-hahsela'-ke
DU-lamp/light-DU
'two lamps,lights' (1973: 235)

{N,NUM,NSG}

wísk nika-hahsela'-ke
five PL-lamp/light-PL
'five lamps/lights' (1973: 236)

4. Restrictions on the patterns in (3) above: *No known restrictions*

² Marianne Mithun is Professor of Linguistics at the University of California, Santa Barbara, USA.
 Email: mithun@linguistics.ucsb.edu

Number systems

- 5. Number distinction: *SG/DU/PL (1973: 234-8)*
- 6. Obligatoriness of number marking: *No information*
- 7. Non-numeral quantifiers: -

Classifier systems

- 8. Noun classes: *Present (1973: 237-238)*
- 9. Numeral classifiers: *No information*

Language: Mongol (Khamnigan)
Family/Genus: Altaic/Mongolic
Country/ Macro Area: China,Mongolia/Eurasia
Reference(s): Janhunen, J. (2003) and Dmitri Morenkov (personal communication)*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

nege *kuun*
one *person*
'one person' (2003: 90)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (2003: 90)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

gurban *kuun*
three *person*
'three persons' (2003: 90)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (2003: 90)*

Number systems

- 5. Number distinction: *SG/PL (2003: 89)*
- 6. Obligatoriness of number marking: *Optional (2003: 89)*
- 7. Non-numeral quantifiers: -

Classifier systems

- 8. Noun classes: *Absent (Dmitri Morenkov, p.c.)*
- 9. Numeral classifiers: *No information*

* Dmitri Morenkov is an Honours M.A. graduate in Mongolian studies, St. Petersburg State University, Russia. Email: tarbagan@inbox.ru.

Language: Mosetén
Family/Genus: Mosetenan/Mosetenan
Country/ Macro Area: Bolivia /South America
Reference(s): Sakel, J. (2004)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

<i>jiri-ty</i>	<i>kojti</i>	<i>daer-tyi'</i>
<i>one-M</i>	<i>heart</i>	<i>big-L.M</i>

'one big heart' (2004: 99)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>paerae'</i>	<i>tara'tara'</i>
<i>two</i>	<i>big.rat</i>

'two big rats' (2004: 84)

{N,NUM,NSG}

<i>Paerae-ki'</i>	<i>tyak</i>	<i>jiri-s</i>	<i>ji-yi-'</i>	<i>soñi'-in</i>
<i>two-size</i>	<i>ten</i>	<i>one-F</i>	<i>PASS-VY-F.S</i>	<i>man-PL</i>

'twenty-one men' (2004: 169)

4. Restrictions on the patterns in (3) above:

{N,NUM}: *No known restrictions*

{N,NUM,NSG}: *Used only with human nouns* (2004: 85)

Number systems

5. Number distinction: *SG/PL* (2004: 81)

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *Reduplication, e.g. jedye' (SG)/jedye'-jedye'(PL) 'thing'*

Classifier systems

8. Noun classes: *Present* (2004: .85-87)

9. Numeral classifiers: *Absent* (WALS)

Language: Mpur
Family/Genus: West Papuan/Kebar
Country/ Macro Area: Indonesia/Australia-New Guinea
Reference(s): Odé, C. (2002)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:
{*N, NUM, CLF*}
jan bik denur
house CLF three
'three houses' (2002: 83)
4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No information*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: *LQ, e.g. fon 'many', as in jan mafun fon [house beautiful many] 'many beautiful houses' (2002: 63)*

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *Present (2002: 83)*

Language: Mulao
Family/Genus: Tai-Kadai/Kam-Tai
Country/ Macro Area: China/Southeast Asia
Reference(s): Guoqiao, Z. (1990)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

nə⁵ kwən⁵ pət⁷

one CLF pen

'one pen' (1990: 173)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

ya² lak⁷ tən²

two CLF cow

'two cows' (1990: 175)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction (1990: passim)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (1990: passim)*

9. Numeral classifiers: *Present (1990: passim)*

Language: Mundang
Family/Genus: Niger-Congo/ Adamawa-Ubangian
Country/ Macro Area: Chad /Africa
Reference(s): Elders, S. (2000)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>yáy-rā</i>	<i>sāy</i>	<i>nò</i>	<i>hīē</i>
<i>maison-PL</i>	<i>trios</i>	<i>CLE</i>	<i>ici</i>
<i>[house-PL]</i>	<i>[three]</i>	<i>[CLE]</i>	<i>here</i>

'Trois maisons se trovent ici.' [Three houses are located here, English translation mine.]
 (2000: 275)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2000: 118)*
6. Obligatoriness of number marking: *Obligatory (WALS)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (2000: 120)*
9. Numeral classifiers: *No information*

Language: Mundari
Family/Genus: Austro-Asiatic/Munda
Country/ Macro Area: India/Eurasia
Reference(s): Sinha, N.K. (1975)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

mid *hoꝛo*

one *man*

'one man' (1975: 111)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

bar *daru-kiꝝ*

two *tree-DU*

'two trees' (1975: 111)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL (NB: only animate nouns) (1975: 60)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (WALS)*

9. Numeral classifiers: *No information*

Language: Nabak
Family/Genus: Trans-New Guinea/Finisterre-Huon
Country/ Macro Area: Papua New Guinea/Australia-New Guinea
Reference(s): Fabian, G., Fabian, E. and Waters, B. (1998)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

<i>bo</i>	<i>sobmaŋ</i>	<i>tembe</i>	<i>kwep</i>	<i>ke</i>
<i>pig</i>	<i>black</i>	<i>big</i>	<i>one</i>	<i>that</i>

'that one big black pig' (1998: 72)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNCNSG:

{N,NUM}

<i>nemba</i>	<i>isisikŋaŋ</i>	<i>zut-aŋ</i>
<i>child</i>	<i>some.little</i>	<i>two-FOC</i>

'two small children' (1998: 148)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1998: 23-24)*

6. Obligatoriness of number marking: *Optional (1998: 23-24)*

7. Non-numeral quantifiers: *LQ, e.g. sambe 'many', as in bo-ŋaŋ sambe [pig-3SG.POSS many] 'his many pigs' (1998: 73)*

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Nahali
Family/Genus: Nahali/Nahali
Country/ Macro Area: India/Eurasia
Reference(s): Shafer, R. (1941)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

bidī awalkā mānchho

a/one good man

'a/one good man' (1941: 363)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNCNSG:

{N,NUM,NSG}

ir awalkā mānchh ā

two good man.PL

'two good men' (1941: 363)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1941: passim)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (1941: passim)*

9. Numeral classifiers: *No information*

Language: Nahuatl (Huasteca)
Family/Genus: Uto-Aztecan/Aztecan
Country/ Macro Area: Mexico /North America
Reference(s): Beller, R. and Beller, P. (1979) and Judith M. Maxwell (personal communication)*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

seh pastiya

one pill

'one pill' (1979: 253)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNCNSG:

{N,NUM,NSG}

eyi tlaka-meh

three man-ANIM.PL

'three men' (1979: 255)

{N,NUM}

ne eyi nelia yehyek-ci I-kali-wa José

those three really pretty-ADJZ his-house-POSSD Joe

'those three really pretty houses of Joe' (1979: 234)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1979: 255)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (Judith M. Maxwell, p.c.).*

9. Numeral classifiers: *Absent (NB: "...though the numbers have reduced forms that can combine with measure word stems and roots to form words which are used as counters, so sort of like numeral classifiers omtel ostotl: two-stone cave") (Judith Maxwell, p.c.).*

* Judith M. Maxwell is Associate Professor of Anthropology at Tulane University, USA. Email: maxwell@tulane.edumaxwell@tulane.edu maxwell@tulane.edu

Language: Nalik
Family/Genus: Austronesian/Oceanic
Country/ Macro Area: Papua New Guinea /Southeast Asia & Oceania
Reference(s): Volker, C. A. (1998) and personal communication*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

a rate be azaxei
 ART man only one
 'only one man' (1998: 117)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNCNSG:

{N,NUM}

a nalik uru
 ART boy two
 'the two boys' (1998: 104)

{N,NSG}

a uru nalik
 ART DU boy
 'the two boys' (1998: 104)

{N,NUM,NSG}

a uban faal orol
 ART PAU house three
 'the three houses' (1998: 106)

4. Restrictions on the patterns in (3) above:

{N,NUM}: *No known restrictions*

{N,NSG}: *No known restrictions*

{N,NUM,NSG}: *Used with non-human nouns (1998: 106)*

Number systems

5. Number distinction: *SG/DU/PAU/PL (1998: 102)*

6. Obligatoriness of number marking: *Obligatory if without numerals; Optional if with numerals (1998: 101)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (Craig A. Volker, p.c.)*

9. Numeral classifiers: *No information*

* Craig A. Volker is Professor of Languages at Gifu Shotoku Gakuen University, Japan. Email: volker@nalik.org

Language: Ndjébbana
Family/Genus: Australia/ Ndjébbana
Country/ Macro Area: Australia (Northern Territory)/Australia-New Guinea
Reference(s): McKay, G. (2000)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,SG}

kayókaya *warábba-na*

night *one-3MIN.M*

'one night' (2000: 294) (*MIN* used as *SG*, 2000:192)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNCNSG

{N,NUM}

nganabbárru *karnayédjabba*

buffalo *two*

'two buffaloes' (2000: 294)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL* (2000: 193)

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ*, e.g. *karrówa* 'many', as in *yídja barra-karrówa* [*man 3AUG-many*] 'many men' (2000: 193)

Classifier systems

8. Noun classes: *Present* (2000: 192)

9. Numeral classifiers: *Absent* (2000: *passim*)

Language: Ndyuka
Family/Genus: Creoles and Pidgins
Country/ Macro Area: French Guiana, Suriname/South America
Reference(s): Huttar, G. L. and M. Huttar (1994)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

wan kodo sama

one only person

'just one person' (1994: 209)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (1994: 74, 209)*

3. Structural patterns of CNNCNSG:

{N,NUM}

tin pikin

ten child

'ten children' (1994: 533)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (1994: 533)*

Number systems

5. Number distinction: *No Distinction (1994: 452)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: *LQ, e.g. somen 'many', as in somen nen [many name] 'some names' (1994: 206)*

Classifier systems

8. Noun classes: *Absent (1994: 452)*

9. Numeral classifiers: *Absent (WALS)*

Language: Nepali
Family/Genus: Indo-European/Indic
Country/ Macro Area: Nepal/Eurasia
Reference(s): Turnbull, A (1923) and Hutt, M. and Subedi, A. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

ek-janā *manche*

one-CLF *person*

'one person' (2003: 49)

{NUM,CLF}

ek *chin*

one *moment*

'one moment' (2003: 56)

2. Restrictions on the patterns in (1) above:

{N,NUM,CLF}: *Used generally with no known restrictions (2003: 49, passim)*

{NUM,CLF}: *Used mostly with nouns denoting measures (2003:56), hence @*

3. Structural patterns of CNNCNSG:

{N,NUM,CLF}

duiṭā *mec*

two.CLF *chair*

'two chairs' (2003: 51) (*vaṭā* =CLF for hon-human nouns)

{NUM,CLF}

cār *din*

four *day*

'four days' (2003: 56)

4. Restrictions on the patterns in (3) above: *Same as (2)*

Number systems

5. Number distinction: *SG/PL (1923: 12)*

6. Obligatoriness of number marking: *Optional (WALS)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Present (2003: passim)*

Language: Nez Perce
Family/Genus: Penutian/Sahaptian
Country/ Macro Area: United States (Idaho, Oregon, Washington)/North America
Reference(s): Rude, N. E. (1985)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNCNSG:

{N,NSG}

'asqáp-iin

brother-DU

'two brothers' (Phinney 1934: 134 in 1985: 77)

{N,NUM.CLF,NSG}

páax-loo' iwéepne-me

five-HUM wife-PL

'five wives' (Phinney 1934: 234 in 1985: 81)

{N,NUM,CLF}

le'épti-t waχ náaqc wa wáa tam

twenty-NONHUM and one.NONHUM trout

'twenty one trout' (Aoki 1970: 138 in 1985: 81) (see (9))

4. Restrictions on the patterns in (3) above:

{N,NSG}: *No known restrictions*

{N,NUM,CLF}: *No known restrictions*

{N,NUM,CLF,NSG}: *Used with human nouns (1985: 76)*

Number systems

5. Number distinction: *SG/DU/PL (1985: 76)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*
9. Numeral classifiers: *Present (numerals are distinguished based on human vs non-human) (1985: 81)*

Language: Ngiyambaa
Family/Genus: Australian/Pama-Nyungan
Country/ Macro Area: Australia (New South Wales)/Australia-New Guinea
Reference(s): Donaldson, T. (1980)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

magu: *miri*

one *dog*

'one dog' (1980: 99)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNCNSG

{N,NUM}

bulagar *miri*

two *dog*

'two dogs' (1980: 102)

{N,DU}

miri-bula:

dog-DU

'a pair of dogs' (1980: 102)

4. Restrictions on the patterns in (3) above

{N,NUM}: *No known restrictions*

{N,DU}: *Used with the implication of 'in a group of two' (1980: 102), i.e. used with the special meaning, hence @.*

Number systems

5. Number distinction: *SG/DU/PL (1980: 102, passim; WALS)*

6. Obligatoriness of number marking: *Optional (WALS)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Absent (WALS)*

Language: Nicobarese (Car)
Family/Genus: Austro-Asiatic/Nicobarese
Country/ Macro Area: India (Nicobar Islands)/Southeast Asia & Oceania
Reference(s): Braine, J. C. (1970)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

héŋ *kuk*

one *coconut*

'one coconut' (1970: 113)

{N,NUM,CLF}

héŋ *nɔŋ* *ʔá:p*

one *CLF* *canoe*

'one canoe' (1970: 113)

2. Restrictions on the patterns in (1) above: *No known restrictions*
3. Structural patterns of CNNC_{NSG}: *No information*
4. Restrictions on the patterns in (3) above: *No information*

Number systems

5. Number distinction: *No distinction (WALS)*
6. Obligatoriness of number marking: *See (5)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (WALS)*
9. Numeral classifiers: *Optional (1970: 50)*

Language: Nigerian Pidgin
Family/Genus: Creoles and Pidgins
Country/ Macro Area: Nigeria/Africa
Reference(s): Faraclas, N. G. (1996)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

wɔn buk

one book

'one book' (1996: 114)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (1996: 114, 261)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

fifti nyam dèm

fifty yam PL

'fifty yams' (1996: 71)

{N,NUM}

fayv kotingrâs

five grass.cutter

'five grass cutters' (1996: 71)

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG}: *Used generally with no known restrictions (1996: 71)*

{N,NUM}: *Used generally with no known restrictions (1996: 71, 169)*

Number systems

5. Number distinction: *SG/PL (1996: 168)*

6. Obligatoriness of number marking: *Optional (1996: 168)*

7. Non-numeral quantifiers: *LQ, e.g. plenti 'be plenty', as in got-got plenti [goat-RDP plenty] 'plenty of goats' (1996: 168); Reduplication, e.g. got (SG)/got-got (PL) 'goats' (1996: 168)*

Classifier systems

8. Noun classes: *Absent (1996: 171)*

9. Numeral classifiers: *Absent (1996: passim)*

Language: Nimboran
Family/Genus: Nimboran/Nimboran
Country/ Macro Area: Indonesia/Australia-New Guinea
Reference(s): May, K. (1997)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

yamo tendʉ

huse one

'one house' (1997: 50)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}

{N,NUM}

hlu ulaŋ namwan

man tall two

'two tall men' (1997: 50)

4. Restrictions on the patterns in (3) above *No known restrictions*

Number systems

5. Number distinction: *No information*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *Reduplication, e.g. yano (SG)/yano-yano (PL) 'village'*

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Nivkh
Family/Genus: Nivkh/Nivkh
Country/ Macro Area: Russia (Siberia)/Eurasia
Reference(s): Gruzdeva, E. (1998)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

<i>n'ivyŋ</i>	<i>n'enŋ</i>
<i>man</i>	<i>one.CLF</i>

'one man' (1998: 63)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG} :

{N,NUM,CLF}

<i>n'ivyŋ</i>	<i>menŋ</i>
<i>man</i>	<i>two.CLF</i>

'two men' (1998: 62)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1998: passim)*
6. Obligatoriness of number marking: *Optional (WALS)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*
9. Numeral classifiers: *Obligatory (1998: 63, passim)*

Language: Norwegian
Family/Genus: Indo-European/Germanic
Country/ Macro Area: Norway/Eurasia
Reference(s): Centre for Information on Language Teaching and Research. (1986) and Kinn, T. (2004)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

én kake

one cake

'one cake' (1986: 148)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (1986: 148)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

fem hundre menn

five hundred.PL man.PL

'five hundred men' (2004: 1)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (2004: 1; 1986: 65)*

Number systems

5. Number distinction: *SG/PL (1986: 64)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ, e.g. mange 'many', as in mange dager [many day-PL] 'many days' (1986: 65)*

Classifier systems

8. Noun classes: *Present (1986: 57)*

9. Numeral classifiers: *Absent (2004: passim)*

Language: Nubi
Family/Genus: Creoles and Pidgins
Country/ Macro Area: Uganda/Africa
Reference(s): Wellens, I.H.W. (2003), and Bernd Heine (personal communication)*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

'wakhid 'zol
one person
'one person' (2003: 90)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

ku'baya ti'nin
cup two
'two cups'

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2003: 58)*

6. Obligatoriness of number marking: *Optional (2003: 58)*

7. Non-numeral quantifiers: *LQ, e.g. lakáta milán [tree many] 'many trees' (Bernd Heine, p.c.)*

Classifier systems

8. Noun classes: *Absent (Bernd Heine, p.c.)*

9. Numeral classifiers: *No information*

* Bernd Heine is Emeritus Professor of Linguistics at the University of Cologne, Germany. Email: bernd.heine@uni-koeln.de

Language: Nuuchahnulth (Nootka)
Family/Genus: Wakashan/Southern Wakashan
Country/ Macro Area: Canada/North America
Reference(s): Aikhenvald, A. (2000), Davidson, M. (2002) and Nichols, J. (1992)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

ʔaʔasuba *ʔa'xuk-badaɣ*

eight *man-PL*

'a group of eight men' (2002: 206)

{N,NUM}

ʔisaʔ *ʔaʔa*

blanket *two*

'two blankets' (2002: 337)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2002: 206)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (1992: 299)*
9. Numeral classifiers: *Absent (2003: 249)*

Language: Ocuilteco
Family/Genus: Oto-Manguenan/Otomian
Country/ Macro Area: Mexico/North America
Reference(s): Muntzel, M. C. (1986/2003)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

m_laa *li_ht*

a/one *f_ish*

'*a/one fish*' (1986: 137)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

Kasuela *m_{ph}yu*

three *bowl*

'*three bowls*' (1986: 94)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL* (1986: 79)

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ*, e.g. *p_ɨkhi* 'many', as in *p_ɨkhi lithaa* [*many-bird*] 'many birds' (1986: 79)

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Olo
Family/Genus: Torricelli/Wapei-Palei
Country/ Macro Area: Papua New Guinea/Australia-New Guinea
Reference(s): Staley, W.E. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:
{*N, NUM, NSG*}

<i>mete</i>	<i>winges</i>
<i>man.PL</i>	<i>two.M</i>

'two men' (2003: 30)
4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (WALS)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: *No information*

Classifier systems

8. Noun classes: *Present (2003: passim)*
9. Numeral classifiers: *No information*

Language: Olutec
Family/Genus: Mixe-Zoque/Mixe-Zoque
Country/ Macro Area: Mexico/North America
Reference(s): Zavala, R. (2002)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>metzo</i>	<i>kawa:yu</i>
<i>two</i>	<i>horse</i>
<i>'two horses' (2002: 480)</i>	

{N,NUM,NSG}

<i>metzko</i>	<i>xujta:tu-tük</i>
<i>two</i>	<i>soldier-PL</i>
<i>'two soldiers' (2002: .63)</i>	

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2002: 41)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: *Third-person plural suffix cross-reference marked on verbs (2002: 42)*

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *No information*

Language: Orig
Family/Genus: Niger-Congo/Kordofanian
Country/ Macro Area: Sudan /Africa
Reference(s): Schadeberg, T. C. and Elias, P. (1979)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,SG}

k-iráj *k-àttá*

CL-thing *CL-one*

'one thing' (1979: 43) (Class k- = Singular number, cf. 1979: 41)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}: *No information*

4. Restrictions on the patterns in (3) above: *No information*

Number systems

5. Number distinction: *SG/PL (1979: 41)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (1979: 41)*

9. Numeral classifiers: *Absent (WALS)*

Language: Pame
Family/Genus: Oto-Manguean/Pamean
Country/ Macro Area: Mexico/North America
Reference(s): Manrique Castaneda, L. (1967)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

nada kuhú
one sorcerer
 'one sorcerer' (1967: 345)

{N,NUM,SG}

nada n-tao
one SG-eye
 'one eye' (1967: 346)

{N,SG}

n-ɸi°
SG-tooth
 'tooth' (1967: 345)

{N,SG}

na-ɸi°
SG-tooth
 'tooth' (1967: 345)

2. Restrictions on the patterns in (1) above:

{N,NUM}: Used generally with no known restrictions (1967: 345)

{N,NUM,SG},{N,SG}: No known restrictions (NB: {N,NUM} unmarked (1967: 345)), hence @

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

pyɛ kudú
four stone
 'four stones' (1967: 346)

{N,NUM,NSG}

tii dyao
two eye.PL
 'two eyes' (1967: 342, 346)

4. Restrictions on the patterns in (3) above: No known restrictions

Number systems

5. Number distinction: SG/PL (1967: 345)

6. Obligatoriness of number marking: No information

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Absent (1967: passim)*

Language: Pech
Family/Genus: Chibchan/Paya
Country/ Macro Area: Honduras /South America
Reference(s): Holt, D. (1999a)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

kàʔó-s

house-one

'a/one house' (1999a: 63)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

šúšú sè:wa òki

dog yellow five

'five yellow dogs' (1999a: 63)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: SG/PL (NB: "Nouns are not normally marked for plurality, but noun subjects can be recognized as plural through plural subject-markers in the verbal complex". (1999a: 38)

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Persian
Family/Genus: Indo-European/ Iranian
Country/ Macro Area: Iran /Eurasia
Reference(s): Mahootian, S. (1997) and Shirin Abadikhah (personal communication)*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

ye(k)-dune hælge

one-CLF ring

'one ring' (Shirin Abadikhah, p.c.)

{N,NUM,SG}

yek hælge-i

one ring-SG

'one ring' (Shirin Abadikhah, p.c.)

{N,SG}

hælge-i

ring-SG

'one ring' (Shirin Abadikhah, p.c.)

{N,NUM}

yek hælge

one ring

'one ring' (Shirin Abadikhah, p.c.)

2. Restrictions on the patterns in (1) above:

{N,NUM,CLF},{N,SG}: No known restrictions but less common than {N,NUM} (Shirin Abadikhah, p.c.), hence @

{N,NUM,SG}: Preferred in literary works rather than in every day use (Shirin Abadikhah, p.c.), hence @

{N,NUM}: No restrictions (Shirin Abadikhah, p.c.)

3. Structural patterns of CNNC_{NSG}:

{N,NUM,(CLF)}

bist (-jeld) ketab

twenty-(CLF) book

'twenty books' (1997: 195)

* Shirin Abadikhah is a native speaker of Persian, studying Linguistics at the University of Edinburgh, UK. Email: s.abadikhah@sms.ed.ac.uk

{N,NUM,(CLF),NSG}

do (-ta) pesær-a

two (-CLF) boy-PL

'the two boys' (1997: 195)

4. Restrictions on the patterns in (3) above:

{N,NUM, (CLF)}: Used with nonspecific nouns (1997: 193)

{N,NUM,(CLF),NSG}: Used with specific nouns (1997:193)

Number systems

5. Number distinction: *SG/PL* (1997: 193)

6. Obligatoriness of number marking: *Obligatory* (NB: *obligatorily unmarked on nonspecific nouns*) (1997: 193)

7. Non-numeral quantifiers: *LQ*, e.g. *un cænd-ta goldun* [*that many-CLF vase*] 'those few vases' (1997: 14)

Classifier systems

8. Noun classes: *Absent* (WALS)

9. Numeral classifiers: *Optional* (1997: 195.)

Language: Piaroa
Family/Genus: Sáliban/Sáliban
Country/ Macro Area: Venezuela /South America
Reference(s): Krute, L.D. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

<i>kurod-æ</i>	<i>hid-æ-tetæ</i>
<i>machete-CLF</i>	<i>one-CLF-one</i>

'one machete' (2003: 144)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF,(NSG)}

<i>kurod(-iy)-æ</i>	<i>wabod-æ-tuhk^wæ</i>
<i>machete(-PL)-CLF</i>	<i>three-CLF-three</i>

'three machetes' (2003: 97, 144)

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF}: *No known restrictions*

{N,NUM,CLF,NSG}: *Rare (2003: 97), hence @*

Number systems

5. Number distinction: *SG/PL (2003: passim)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ, e.g. kurodæ reh^wæodæ [many.CLF machete] 'many machetes' (2003: 97)*

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Present (2003: 97, passim)*

Language: Pilagá
Family/Genus: Guaicuruan /Guaicuruan
Country/ Macro Area: Argentina/South America
Reference(s): Vidal, A. (2002)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>qanač'e</i>	<i>na'</i>	<i>tayiñi</i>	<i>dosol-qa</i>	<i>na'</i>	<i>emek-qa</i>
<i>conj</i>	<i>DEICLF³</i>	<i>south</i>	<i>two-PAU</i>	<i>DEICLF</i>	<i>house-PAU</i>

'And in direction to the south, there are two houses.' (2002: 129)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PAU/PL (PL = COLL) (2002: 91-96)*
6. Obligatoriness of number marking: *Obligatory (WALS)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (NB: not productive) (2002: 11)*
9. Numeral classifiers: *No information*

³ Deitic classifiers “categorize the noun in terms of its shape, animacy, and position in space” (Aikhenvald 2000: 176)

Language: Pirahã
Family/Genus: Mura/Mura
Country/ Macro Area: Brazil/South America
Reference(s): Everett, D. (2005)

CNNCs

1. Structural patterns of CNNC_{SG}: *No numeral 'one' proper, hence no CNNC_{SG}*

a. tí 'ítí'isi hóí hii 'aba'áígio 'oogabagáí
 ISG fish small PRED only want
 'I only want {one/a couple/a small} fish.' (2005: 623)

NB: "This could not be used to express a desire for one fish that was very large, except as a joke". (NB original)

b. tiobáhai hóí hii
 child small PRED
 'small child/child is small/one child' (2005: 623)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (2005: passim)*

3. Structural patterns of CNNC_{NSG}: *No numerals proper, hence no CNNC_{NSG}*

4. Restrictions on the patterns in (3) above: *See (3)*

Number systems

5. Number distinction: *No distinction (2005: 623)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: *LQ, hóí 'larger', e.g.*

tí 'ítí'isi hóí hii 'oogabagáí
 ISG fish larger PRED want
 'I want [a few/larger/small] fish.' (2005: 623)

Classifier systems

8. Noun classes: *Present (WALS)*

9. Numeral classifiers: *Absent (WALS)*

Language: Qiang
Family/Genus: Sino-Tibetan/Qiangic
Country/ Macro Area: China/Southeast Asia & Oceania
Reference(s): Lapolla, R. J. (2003c) and LaPolla, R. J. (with C. Huang) (2003d)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

khuə e-ze

dog one-CLF

'one dog' (2003b: 67)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions* (2003c: 580; 2003d: 67)

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

ləʔz-jə-pen

book-two-CLF

'two books' (2003c: 82)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions* (2003c: 579; 2003d: 82)

Number systems

5. Number distinction: *SG/PL* (2003c: 69)

6. Obligatoriness of number marking: *Optional* (2003c: 69)

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent* (2003c, 2003d: *passim*)

9. Numeral classifiers: *Present* (2003c, 2003d: *passim*)

Language: Quechua (Huallaga)
Family/Genus: Quechuan/Quechuan
Country/ Macro Area: Peru/South America
Reference(s): Nichols, J. (1992) and Weber, D. J. (1989)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>Chay</i>	<i>kapilla</i>	<i>ishakay</i>	<i>punku-yoq</i>
<i>that</i>	<i>chapel</i>	<i>two</i>	<i>door-have</i>

'That chapel has two doors.'

{N,NUM,NSG}

<i>chay</i>	<i>ishkay</i>	<i>hatun</i>	<i>wasi-kuna</i>
<i>that</i>	<i>two</i>	<i>big</i>	<i>house-PL</i>

'those two big houses.' (1989: 17)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL* (1989: 55)
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: *LQ*, e.g. *achka* 'many', as in *achka wata* 'many years' (1989: 18)

Classifier systems

8. Noun classes: *Absent* (1992: 301)
9. Numeral classifiers: *Absent* (WALS)

Language: Quileute
Family/Genus: Chimakuan/Chimakuan
Country/ Macro Area: United States (Washington)/North America
Reference(s): Andrade, M. J. (1933), Nichols, J. (1992)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

tcik ^u	ka·yad	ɬawé··k'wa'-as	hahé't'c
big	shark	two-daughter-SBJ.3SG	pretty-REDUP

'Big shark had two daughters who were pretty.' (1933: 191)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No information*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: *Reduplication, e.g. da'q'ò (SG)/dadá'q'o (PL) 'eye' (1933)*

Classifier systems

8. Noun classes: *Present (1992: 299)*
9. Numeral classifiers: *Absent (WALS)*

Language: Russian
Family/Genus: Indo-European/Slavic
Country/ Macro Area: Russia/Eurasia
Reference(s): Neidle, C. (1988)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

odna studentka
one.F student.F
 'one student' (1988: 102)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,OBL,SG}

dva dnja
two day.GEN.SG
 'two days' (1988: 95)

{N,NUM,OBL,NSG}

pjat' mal'čikov
five boy.GEN.PL
 'five boys' (1988: 90)

{N,NUM}

dvadcat' odin student
twenty one student
 'twenty-one students' (1988: 102)

4. Restrictions on the patterns in (3) above:

{N,NUM,OBL,SG}: *Used with the numerals 2-4 (1988: 95)*

{N,NUM,OBL,NSG} *Used with the numerals 5 and above, except some numerals such as those ending with 1 where {N,NUM} is used (e.g. 21) (1988: 90-95)*

Number systems

5. Number distinction: *SG/PL (WALS)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: *LQ, e.g. mnogo 'many', as in mnogo mal'čikov [many boy.GEN.PL] 'many boys'*

Classifier systems

8. Noun classes: *Present (WALS)*

9. Numeral classifiers: *Absent (WALS)*

Language: Santali
Family/Genus: Austro-Asiatic/Munda
Country/ Macro Area: India /Eurasia
Reference(s): Campbell, G. L. (2000) and Neukom, L. (2001)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NSG}

sadom-kin

horses-DU

'two horses' (2000: 1454)

{N,NUM}

bar *hor*

two *person*

'two persons' (2000: 1456)

{N,NUM,CLF}

m̃r̃ē-gɔɛn *əiməl*

five-CLF *woman*

'five women' (2001: 55)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL (2001: 21)*
6. Obligatoriness of number marking: *Optional (2001: 21)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *Optional (WALS)*

Language: Sanuma
Family/Genus: Yanomam/Yanomam
Country/ Macro Area: Brazil, Venezuela/South America
Reference(s): Borgman, D. M. (1990)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

<i>sami-i</i>	<i>ipa</i>	<i>po</i>	<i>piho</i>
<i>one-INDEF</i>	<i>my</i>	<i>machete</i>	<i>give</i>

'give me one machete.' (1990: 140)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>tute-i</i>	<i>ipa</i>	<i>polakapi-i</i>	<i>ipa</i>	<i>po</i>
<i>new-INDEF</i>	<i>1SG.POSS</i>	<i>two-INDEF</i>	<i>1SG.POSS</i>	<i>machete</i>

<i>kökö</i>	<i>hanaha</i>	<i>pi</i>	<i>-ta</i>
<i>3DU</i>	<i>red</i>	<i>give</i>	<i>-EXT</i>

'Give me my new, two red machetes.' (1990: 141)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (NB: through classifier pronouns (see (3)) (1990: 141)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Sapuan
Family/Genus: Austro-Asiatic/Bahnaric
Country/ Macro Area: Laos/Southeast Asia & Oceania
Reference(s): Jacq, P. and Sidwell, P. (1999) and Paul Sidwell (personal communication)*

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:
{N, NUM, CLF}
kuan tham ra
child eight CLF (lit. 'person')
'eight children' (1999: 17)
4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction (Paul Sidwell, p.c.)*
6. Obligatoriness of number marking: *See (5)*
7. Non-numeral quantifiers: *LQ, e.g. ?ɛn 'many', as in mu ?ɛn ra [friend many CLF] 'many friends' (1999: 17)*

Classifier systems

8. Noun classes: *Absent (Paul Sidwell, p.c.)*
9. Numeral classifiers: *Present (NB: impoverished) (Paul Sidwell, p.c.)*

* Paul Sidwell is Director of Mon-Khmer Languages Project and Visiting Fellow of Australian National University, Australia. Email: paulsidwell@yahoo.com

Language: Sarcee
Family/Genus: Na-Dene/Athapaskan
Country/ Macro Area: Canada /North America
Reference(s): Cook, E. (1984)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>gū̀nìsnóni-ká</i>	<i>tí-ká</i>
<i>nine-PL</i>	<i>dog-PL</i>
<i>'nine dogs'</i>	

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG}: *Note that nouns are not normally marked for number (1984: 65), hence {N,NUM,NSG} is not a primary mode of CNNC_{NSG}.*

Number systems

5. Number distinction: *SG/PL (NB: only for kinship terms and some animals) (1984: 65)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *No information*

Language: Selkup
Family/Genus: Uralic/Samoyedic
Country/ Macro Area: Russia/Eurasia
Reference(s): Helimski, E. (1998)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

ukkir qum

one person

'one person' (1998: 575)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

šitti qum

two person

'two persons' (1998: 575)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (WALS)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (1998: passim)*

9. Numeral classifiers: *Absent (1998: passim)*

Language: Semelai
Family/Genus: Austro-Asiatic/Aslian
Country/ Macro Area: Malaysia/Southeast Asia & Oceania
Reference(s): Kruspe, N. (2004)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

mə=ʔikur cɔ

one=CLF dog

'one dog' (2004: 99)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

mpɔh bjɛʔ c^hɔŋ

seven CLF hill

'seven hills' (2004: 204)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No information*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Obligatory (WALS)*

Language: Sentani
Family/Genus: Sentani/Sentani
Country/ Macro Area: Indonesia /Australia-New Guinea
Reference(s): Cowan, H.K.J. (1965)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

ja əmbaj

day one

'one day, a certain day' (1965: 58)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

do namε

man four

'four men' (1965: 58)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction (1965: passim)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *No information*

Language: Shabo
Family/Genus: Nilo-Saharan/Shabo
Country/ Macro Area: Ethiopia/Africa
Reference(s): Teferra, A. (1989)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}

{*N, NUM*}

bap mat c'iiij kaan
two big black dog

'two big black dogs' (1989: 382)

4. Restrictions on the patterns in (3) above *No known restrictions*

Number systems

5. Number distinction: *SG/PL (WALS)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *No information*

Language: Somali
Family/Genus: Afro-Asiatic/Eastern Cushitic
Country/ Macro Area: Somalia /Africa
Reference(s): Bell, C.R.V. (1953), Saeed, J. I. (1993) and (1999)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,UNIT}

<i>xabbàd</i>	<i>lín</i>	<i>áh</i>
<i>unit</i>	<i>orange</i>	<i>be</i>

'one orange' (lit. 'a unit that is orange') (1999: 58)

{N,NUM}

<i>hál</i>	<i>baabúur</i>
<i>one</i>	<i>truck</i>

'one truck' (1999: 70)

2. Restrictions on the patterns in (1) above:

{N,UNIT}: Used with only some nouns which are transnumeral (1999: 58), hence @ (NB: This structural pattern is not mentioned at all in other Somali reference grammars, so the structural pattern is regarded here as non-primary)

{N,NUM}: No known restrictions

3. Structural patterns of CNNC_{NSG}

{N,NUM}

<i>shan</i>	<i>nin</i>
<i>five</i>	<i>man</i>

'five man' (1953: 49)

{N,NUM,OBL,SG}

<i>sàddex</i>	<i>wíl</i>
<i>three</i>	<i>boy.GEN</i>

'three boys' (lit. three of boy) (1999: 71)

{N,NUM,OBL,NSG}

<i>afar</i>	<i>naagóod</i>
<i>four.ABS</i>	<i>woman.GEN.PL</i>

'four women' (lit. 'four of women) (1999: 70)

{N,NUM,NSG}

<i>shan</i>	<i>lah-aad</i>
<i>five</i>	<i>sheep-PL</i>

'five sheep' (1953: 49)

4. Restrictions on the patterns in (3) above

{N,NUM}: Used generally with no known restrictions (1953: 49)

{N,NUM,OBL,SG}: No known restrictions

{N,NUM,OBL,NSG}: No known restrictions

{N,NUM,NSG}: Used with nouns denoting domestic animals(ending in –aad) (1953: 49), hence @

Number systems

5. Number distinction: SG/PL (1999: 62)

6. Obligatoriness of number marking: *Obligatory* (WALS)

7. Non-numeral quantifiers: *LQ*, e.g. dhàwr, as in dhàwr nín [*several man.GEN*] ‘several men’ (lit. *several of man*) (1993: 183); *Reduplication*, e.g. jid (SG)/jidàd (PL) ‘road’; jür (SG)/jüràr(PL) ‘rat’ (1999:48)

Classifier systems

8. Noun classes: *Present* (1999: 54)

9. Numeral classifiers: *Absent* (1999: *passim*)

Language: Sulka
Family/Genus: Sulka/Sulka
Country/ Macro Area: Papua New Guinea/Australia-New Guinea
Reference(s): Nichols, J. (1992) and Tharp, D. (1996)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,SG}

a-tou *atgiang*

SG.NNI-yam *one*

'one yam' (1996: 115)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

o-sngu *korlolo*

PL.NNI-yam.PL *four*

'four yams' (1996: 115) (sngu = suppletive form of tou 'yam')

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL (1996: passim)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (1992: 298-299)*

9. Numeral classifiers: *Absent (1992: 298-299)*

Language: Taba
Family/Genus: Austronesian/ South Halmahera-West New Guinea
Country/ Macro Area: Indonesia /Southeast Asia & Oceania
Reference(s): Bowden, J. (2001)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

bbuk *p-so*
book *CLF-one*
‘one book’ (2001: 214)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions* (2001: 214, 256)

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

gocila kop sio
corn CLF nine
‘nine grains of corn’ (2001: 242)

{N,NUM,CLF,NSG}

mapin=si mat=tol
woman=PL CLF=three
‘three women’ (2001: 256)

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF}: *Used generally with no known restrictions* (2001: 254, 257)

{N,NUM,CLF,NSG}: *Used with human nouns* (2001: 190-191, 256)

Number systems

5. Number distinction: *SG/PL (NB: only human nouns)* (2001: 190-191)

6. Obligatoriness of number marking: *Obligatory* (2001: 190-191)

7. Non-numeral quantifiers: *LQ, e.g. lloci ‘many’, as in um lloci [house many] ‘many houses’* (2001: 204)

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Obligatory* (2001: 242)

Language: Tamang
Family/Genus: Sino-Tibetan/Bodic
Country/ Macro Area: Nepal/Southeast Asia & Oceania
Reference(s): Mazaudon, M. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

⁴*kle* ⁴*kik*
king *one*
'one king' (2003: 299)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

⁴*came* ⁴*ni:*
daughter *two*
'two daughters' (2003: 294)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction* (2003: 299)

6. Obligatoriness of number marking: *See* (5)

7. Non-numeral quantifiers: *Collective marker, -kate* (2003: 299)

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Absent* (2003: *passim*)

Language: Tamil
Family/Genus: Dravidian/Southern Dravidian
Country/ Macro Area: India, Sri Lanka/Eurasia
Reference(s): Lehmann, T. (1989)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM,(NSG)}

irantu *naay (-kal)*

two *dog(-PL)*

'two dogs come'. (1989: 21)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1989: passim)*
6. Obligatoriness of number marking: *Obligatory, but optional in non-rational nouns (1989: passim)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (WALS)*
9. Numeral classifiers: *Absent (1989: passim)*

Language:	Tariana
Family/Genus:	Arawakan/Arawakan
Country/ Macro Area:	Brazil /South America
Reference(s):	Aikhenvald, A. (2002) and (2003)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

<i>paita</i>	<i>emite</i>
<i>one.CLF.ANIM</i>	<i>youngster.CLF.ANIM</i>
<i>'one child' (2003: 93)</i>	

2. Restrictions on the patterns in (1) above:

{N,NUM,CLF}: *Used generally with no known restrictions (2003: 88, 93, 591)*

3. Structural patterns of CNNC_{NSG}

{N,NUM,CLF,NSG}

<i>kehpuni-dapana-pe</i>	<i>panisi-pe</i>
<i>four-CLF-PL</i>	<i>house-PL</i>
<i>'four houses' (2002:99)</i>	

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF,NSG}: *Used generally with no known restrictions (2002: 99; 2003: 176, 476)*

Number systems

5. Number distinction: *SG/PL* (2003: 165)

6. Obligatoriness of number marking: *Obligatory for animate nouns; Optional for inanimate nouns (2003: 165)*

7. Non-numeral quantifiers: *LQ*, e.g. *hanupe* 'many', as in *itʃida-pe hanupe [turtle-PL many]* 'a lot of turtles' (2003: 220)

Classifier systems

8. Noun classes: *Present* (2003: 87)

9. Numeral classifiers: *Present* (2003: 103, 207)

Language: Teribe
Family/Genus: Chibchan/Talamanca
Country/ Macro Area: Costa Rica, Panama /South America
Reference(s): Quesdada, J. D. (2000)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

sbi kw-ara

pot CLF-one

'one pot' (2000: 11)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

shiii doglo mya

dog CLF three

'three dogs' (2000: 86)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (NB: mostly nouns denoting humans and some animals) (2000: 52, 146)*

6. Obligatoriness of number marking: *Optional (WALS)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *Obligatory (WALS)*

Language: Thai
Family/Genus: Tai-Kadai/Kam-Tai
Country/ Macro Area: Thailand/Southeast Asia & Oceania
Reference(s): Own knowledge (native)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

mǎ: nùŋ tua

dog one tua

'one dog'

{NUM,CLF}

nùŋ wan

one day

'one day'

{N,CLF}

ca:n baj

plate CLF

'one plate'

2. Restrictions on the patterns in (1) above:

{N,NUM,CLF}: Used generally but not with nouns denoting units

{NUM,CLF}: Used with nouns denoting units, such as day, kilometer, hence @

{N,CLF}: Used only in spoken language, hence @

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

mǎ: sǎŋ tua

dog two tua

'two dogs'

{NUM,CLF}

sǎŋ wan

two day

'two days'

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF}: Used generally but not with nouns denoting units

{NUM,CLF}: Used only with nouns denoting units, such as day, kilometer, hence @

Number systems

5. Number distinction: No distinction

6. Obligatoriness of number marking: See (5)

7. Non-numeral quantifiers: *LQ*, e.g. *lǎy* 'several', as in *lǎy wan* [several day] 'several days';
Reduplication, *dèk* (SG)/*dèk-dèk* (PL) 'child'

Classifier systems

8. Noun classes: *Absent*

9. Numeral classifiers: *Obligatory*

Language: Tibetan (Standard Spoken)
Family/Genus: Sino-Tibetan/Bodic
Country/ Macro Area: China (Tibet)/Southeast Asia & Oceania
Reference(s): Denwood, P. (1999)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:
{N,NUM}
deb. Nyi.shu
book twenty
'twenty books' (1999: 101)
4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction (1999: 100)*
6. Obligatoriness of number marking: *See (5)*
7. Non-numeral quantifiers: *LQ, e.g. mang.po 'many', as in deb. mang.po [book a.lot.of] 'a lot of books' (1999: 100-101)*

Classifier systems

8. Noun classes: *Absent (WALS)*
9. Numeral classifiers: *Absent (1999: passim)*

Language: Tidore
Family/Genus: West Papuan/North Halmaheran
Country/ Macro Area: Indonesia/Australia-New Guinea
Reference(s): van Staden, M. (2000)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

nyao delo lamo rimoi

fish tuna big one

'one large tuna fish' (2000: 131)

{N,NUM,CLF}

igo futu-moi

coconut CLF-one

'one coconut tree' (2000: 124)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

ngofa ngarukange

child CLF.three

'three children' (2000: 198)

{N,NUM}

ngofa range

child three

'three children.' (2000: 364)

4. Restrictions on the patterns in (3) above: *No known restrictions NB: "Nowadays, sortal classifiers are always optional, although native speakers consider the forms with the classifier 'much better Tidore'. Only in cases where ambiguity might arise without a classifier, such as the difference between sic coconut trees or six coconuts..., the classifier is invariably present..." (2000: 167)*

Number systems

5. Number distinction: *No distinction (WALS)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: *LQ, e.g. many 'dofu', as in kason dofu 'many dogs' (2000: 164)*

Classifier systems

8. Noun classes: *Present (WALS)*

9. Numeral classifiers: *Optional (2000: 167)*

Language: Tiwi
Family/Genus: Australian/Tiwian
Country/ Macro Area: Australia (Northern Territory)/Australia-New Guinea
Reference(s): Osborne, C. R. (1974)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

jura.ɬa *pəlaŋəmɰawɰi*

two *dog.PL*

'two dogs' (1974: 73)

{N,NUM}

ɬaɬərima *umɰŋa*

three *day*

'three days' (1974: 52)

4. Restrictions on the patterns in (3) above

{N,NUM,NSG}: *Used with human nouns and some animals (e.g. dogs) (1974: 52-53)*

{N,NUM}: *Used with non-human nouns (1974: 52-53)*

Number systems

5. Number distinction: *SG/PL (1974: 52-53)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: *LQ, e.g. taikuwani 'many', as in taikuwani waliwalini [many ant] 'many ants' (1974: 52); Reduplication, e.g. awurini 'man'/wawuruwi 'men' (1974: 53)*

Classifier systems

8. Noun classes: *Present (WALS)*
9. Numeral classifiers: *Absent (WALS)*

Language: Tlingit
Family/Genus: Na-Dene/Tlingit
Country/ Macro Area: United States (Alaska)/North America
Reference(s): Naish, C. M. (1979)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:
{N,NUM}

<i>du</i>	<i>nλs'g</i>	<i>x'úx'u</i>
<i>his</i>	<i>three</i>	<i>book</i>

'his three book' (1979: 103)
4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (WALS)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *Present (WALS)*

Language: Tok Pisin
Family/Genus: Creoles and Pidgins
Country/ Macro Area: Papua New Guinea/Australia-New Guinea
Reference(s): Dutton, T.E and Thomas, D. (1985) and Verhaar, J. W.M. (1995)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

*wan-pela*⁴ *boi*

one-ATTR *boy*

'one boy' (1995: 29)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions* (1985: 67, 207; 1995: 29)

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

tu-kilo

two-kilo

'two kilos' (1995: 169)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions* (1985: 143; 1995: 169)

Number systems

5. Number distinction: *No distinction* (1995: *passim*)

6. Obligatoriness of number marking: *See* (5)

7. Non-numeral quantifiers: *LQ*, e.g. *olgeta* 'all', as in *olgeta dok* [*all-dog*] 'all dogs'; *sam* 'some', as in *sam-pela dok* [*some-ATTR dog*] 'some dogs' (1985: 67-68); *Plural word ol*, e.g. *ol dok* [*dog-PL*] 'dogs' (1985: 67)

Classifier systems

8. Noun classes: *Absent* (1985: *passim*)

9. Numeral classifiers: *Absent* (1985: *passim*)

⁴ "The suffix *-pela* is used to form a limited number of modifiers to nouns, [...] some pronouns [...], and a variety of numerals" (1995: 12).

Language: Tol
Family/Genus: Tol/Tol
Country/ Macro Area: Honduras/North America
Reference(s): Holt, D. (1999b)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>p'is</i>	<i>mát'e-ya</i>	<i>y-us</i>	<i>t^hi-nyúk</i>
<i>deer</i>	<i>two-ADJ</i>	<i>3PL -ACC</i>	<i>PST-see.1SG</i>

'I saw the two deer.' (1999b: 45)

{N,NUM,NSG}

<i>ne-yóm</i>	<i>mát'e-ya</i>
<i>PL-man</i>	<i>two-ADJ</i>

'two men' (1999b: 48)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (NB: only animate nouns)* (1999b: 38)
6. Obligatoriness of number marking: *Optional* (1999b: 38)
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*
9. Numeral classifiers: *Absent (WALS)*

Language: Totonac (Misantla)
Family/Genus: Totonacan/Totonacan
Country/ Macro Area: Mexico/North America
Reference(s): MacKay, C. J. (1991)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:
{*N, NUM, CLF*}
kaak-taat ěik
CLF-four house
'four houses' (1991: 480)
4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1991: 405)*
6. Obligatoriness of number marking: *Optional (1991: 414)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *Optional (1991: passim)*

Language: Trumai
Family/Genus: Trumai/Trumai
Country/Macro Area: Brazil/South America
Reference(s): Guirardello-Damian, R. (1999) and personal communication*

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

huch kiki a yi chetsi
two man DU yi fall

'Two men fell.' (1999: 57) (NB: yi = not glosed)

{N,NUM}

huch tahu
two knife

'two knives' (1999: 23)

4. Restrictions on the patterns in (3) above:
 {N,NUM,NSG}: *No known restrictions*
 {N,NUM}: *Used with inanimate nouns (1999: passim)*

Number systems

5. Number distinction: *SG/DU/PL (1999: passim)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: *Collective paine, "this word is on its way to becoming a grammatical morpheme, because very often its form is reduced to pa, almost forming a phonological unit with the element it modifies: axos paine 'collective of children (i.e. all the children in the village)' axos pa 'collective of children'". (1999: 56)*

Classifier systems

8. Noun classes: *Absent (Raquel Guirardello-Damian, p.c.)*
9. Numeral classifiers: *Absent (Raquel Guirardello-Damian, p.c.)*

* Raquel Guirardello-Damian (PhD) is Visiting Lecturer in Linguistics at the University of the West of England, UK. Email: Raquel.Guirardello-Damian@uwe.ac.uk

Language: Tsimshian (Coast)
Family/Genus: Penutian/Tsimshian
Country/ Macro Area: Canada, United States/North America
Reference(s): Dunn, J. A. (1979)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

k' üül *baa'lx*

one.CLF *ghost*

'one ghost' (1979: 56)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (1979: 38)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF,NSG}

gu'pl *bu-baa'lx*

two.CLF *PL-ghost*

'two ghost' (1979: 56)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (1979: 56)*

Number systems

5. Number distinction: *SG/PL (1979: 13)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *Reduplication, e.g. dasx (SG)/ dikdasx (PL) 'squirrel' (1979: 13)*

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Obligatory (WALS)*

Language: Tucano
Family/Genus: Tucanoan/Tucanoan
Country/ Macro Area: Brazil, Colombia/South America
Reference(s): Barnes, J. (1999)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

<i>ni'cā-wā</i>	<i>wāĩquẽ'a</i>
<i>one-CLF</i>	<i>manioc.squeezer</i>
<i>'one manioc squeezer' (1999: 218)</i>	

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}: *No information*

4. Restrictions on the patterns in (3) above: *No information*

Number systems

5. Number distinction: *SG/PL (1999: passim)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (1999: 221)*

9. Numeral classifiers: *Optional (WALS) (NB: obligatory in some dialects 1999: 219)*

Language: Tukang Besi
Family/Genus: Austronesian/Sulawesi
Country/ Macro Area: Indonesia/Southeast Asia & Oceania
Reference(s): Donohue, M (1999a) and personal communication*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

miapande sa-mia
shaman one-CLF
'one shaman' (1999a: 110)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

sia-rope na kapala
nine-CLF boat.NOM
'nine motor ships' (1999a: 110)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction (Mark Donohue, p.c.)*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: *LQ, e.g. koruo 'many', as in po'o koruo [mango many] 'many mangoes' (1999: 58)*

Classifier systems

8. Noun classes: *Absent (Mark Donohue, p.c.)*

9. Numeral classifiers: *Optional (WALS)*

* Mark Donohue is Professor of Linguistics at the University of Monash, Australia. Email: Mark.Donohue@arts.monash.edu.au

Language: Turkish
Family/Genus: Altaic/Turkic
Country/ Macro Area: Turkey/Eurasia
Reference(s): Kornfilt, J. (1997)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

<i>bir</i>	<i>çürük</i>	<i>elma</i>
<i>one</i>	<i>rotten</i>	<i>apple</i>

'one rotten apple' (1997: 275)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions* (1997: 229, 275)

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>beş</i>	<i>çocuk</i>
<i>five</i>	<i>child</i>

'five children' (1997: 265)

{N,NUM,CLF}

<i>beş</i>	<i>tane</i>	<i>elma</i>
<i>five</i>	<i>item</i>	<i>apple</i>

'five apples' (1997: 271)

4. Restrictions on the patterns in (3) above:

{N,NUM}: *Used generally with no known restrictions* (1997: 229, 275)

{N,NUM,CLF}: *Used restrictively* (1997: 271), hence @

Number systems

5. Number distinction: *SG/PL* (1997: 265)

6. Obligatoriness of number marking: *Obligatory* (1997: 265)

7. Non-numeral quantifiers: *LQ*, e.g. *birçok* 'many', as in *birçok çocuk* [many child] 'many children.' (1997: 266); *bazı çocuk-lar* [some child-PL] 'some children' (1997: 266) (NB: some quantifiers require a quantified noun in plural)

Classifier systems

8. Noun classes: *Absent* (1997: 270)

9. Numeral classifiers: *Present* (NB: not productive) (1997: 271)

Language: Tuvaluan
Family/Genus: Austronesian/Oceanic
Country/ Macro Area: Tuvalu/Southeast Asia & Oceania
Reference(s): Besnier, N. (2000)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>te</i>	<i>tinogafulu</i>	<i>faafine</i>
<i>DEF</i>	<i>ten</i>	<i>woman.PL</i>

'the ten women.' (2000: 361)

{N,NUM,NSG,NUMPCL}

<i>te</i>	<i>toko</i>	<i>tolu</i>	<i>taagata</i>
<i>DEF</i>	<i>NUMPCL</i>	<i>three</i>	<i>man.PL</i>

'the three men' (2000: 360)

{N,NUM}

<i>lua</i>	<i>pii</i>
<i>two</i>	<i>drinking-coconut</i>

'two coconuts' (2000: 360)

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG}: *No known restrictions*

{N,NUM,NSG,NUMPCL}: *No known restrictions*

{N,NUM}: *No known restrictions, but preferred among the three (2000: 360), hence {N,NUM,NSG} and {N,NUM,NSG,NUMPCL} = @*

Number systems

5. Number distinction: *SG/PL (2000: 359)*
6. Obligatoriness of number marking: *Obligatory (2000: 359)*
7. Non-numeral quantifiers: *LQ, e.g. uke 'many', as in (e) toko uke tino [(NPST) NM many person] 'many people' (2000: 576) (NB: "several quantifiers have verb-like properties" (2000: 575))*

Classifier systems

8. Noun classes: *Absent (2000: 367)*
9. Numeral classifiers: *Absent (2000: 367)*

Language: Udihe
Family/Genus: Altaic/Tungusic
Country/ Macro Area: Russia (Siberia)/Eurasia
Reference(s): Nikolaeva, I. A. and Tolskaya, M. (2001)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

omo kusige-we-de xebu-je
 one knife-ACC-FOC take-IMP.2SG
 'Take (at least) one knife.' (2001: 443)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

zu: mafa eme:-ni
 two bear come.PST-3SG
 'Two bears came.' (2001: 537)

{N,NUM,NSG}

za: n̄'aula-ziga omo n̄'aula-tigi w'ali-e-ti
 four boy-PL one boy-LAT fight-PST-3PL
 'four boys fought with one boy.' (2001: 488)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (NB: animate nouns more likely)* (2001: 116)

6. Obligatoriness of number marking: *Optional* (2001: 103)

7. Non-numeral quantifiers: "semantic agreement with the adjectival modifier and/or with the predicate".

(2001: 115), e.g. o-lo bi:-ti tuja ni:
 this-LOC be-3PL five man
 'There are five men here.' (2001: 492).

Classifier systems

8. Noun classes: *Absent* (2001: passim)

9. Numeral classifiers: *Absent* (2001: passim)

Language: Urarina
Family/Genus: Urarina/Urarina
Country/ Macro Area: Peru /South America
Reference(s): Olawsky, K J. (2002) and personal communication*

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

lejhiĩ katça

one man

'one man' (2002: 55)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

kuruwataha-ẽ katça

two-CLF man

'two men' (2002: 49)

{N,NUM,CLF,NSG}

niçataha-ẽ (raj) bere-ẽrũ

three-CLF (3SG.POSS) child-PL

'(his) three children' (2002: 83)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2002: 10)*

6. Obligatoriness of number marking: *Optional (2002: 10)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (Knut J. Olawsky, p.c.)*

9. Numeral classifiers: *Present (2002: 49)*

* Knut J. Olawsky (PhD) was a research fellow at Research Centre for Linguistic Typology, La Trobe University, Australia. Email: olawksy@gmx.net

Language: Urubú-kaapor
Family/Genus: Tupian/Tupi-Guaraní
Country/ Macro Area: Brazil/South America
Reference(s): Kakumasu, J. (1986)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

peteĩ renda

one place

'one place' (1986: 374)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

mokõi pytun

two night

'two nights' (1986: 374)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (WALS)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ, e.g. heta 'many', as in pira riki heta [fish EMPH many] 'there are lots of fish' (1986: 333)*

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *No information*

Language: Usarufa
Family/Genus: Trans-New Guinea/Eastern Highlands
Country/ Macro Area: Papua New Guinea/Australia-New Guinea
Reference(s): Bee, D (1973)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

móra naa-ma

one house-NOM

'one house' (1973: 278)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}: *No information*

4. Restrictions on the patterns in (3) above: *No information*

Number systems

5. Number distinction: *SG/DU/PL (1973: 251)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ, e.g. netuq 'many', as in netuq-wááyúkámá [plenty-man] 'a lot of men' (1973: 304)*

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Vietnamese
Family/Genus: Austro-Asiatic/Viet-Muong
Country/ Macro Area: Vietnam/Southeast Asia & Oceania
Reference(s): Bisang, W. (1999) and Nguyen, T.H. (2004)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

tôi mua một quả cam
1SG buy one CLF orange

'I buy the orange. /I buy an orange.' (1999: 146)

{N,CLF}

tôi mua quả cam
1SG buy CLF orange

'I buy the orange. /I buy an orange.' (Bisang 1996: 541 in 1999: 146)

2. Restrictions on the patterns in (1) above:

{N,NUM,CLF}: No known restrictions

{N,CLF}: Used in a special context "The classifier-noun construction [...] is a special type of a nominal compound which is used in discourse with aforementioned entities which cannot be referred to by pronouns", hence @

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

hai con mèo
two CLF cat

'two cats' (2004: 99)

{NUM,CLF}

bốn ngày
four day

'four days' (2004: 101)

4. Restrictions on the patterns in (3) above:

{N,NUM,CLF}: No known restrictions

{NUM,CLF}: Used with nouns denoting units (2004: 101), hence @

Number systems

5. Number distinction: No information

6. Obligatoriness of number marking: No information

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Obligatory (WALS)*

Language: Wambaya
Family/Genus: Australian/West Barkly
Country/ Macro Area: Australia (Northern Territory)/Australia-New Guinea
Reference(s): Nordlinger, R. (1998)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

<i>gardawuga</i>	<i>ngiy-a</i>	<i>wankurarri</i>	<i>marrgulu</i>
<i>one.IV.ACC</i>	<i>3SG.NM.A-PST</i>	<i>lay</i>	<i>egg.IV.ACC</i>

'She laid one egg.' (1998: 73)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>murrgunji</i>	<i>alaji</i>
<i>three.I</i>	<i>boy.I</i>

'three boys.' (1998: 78)

{N,NSG}

darrangu-wulu
stick-DU
 'two sticks.' (1998: 74)

{N,NUM,NSG}

<i>gujarrawulu</i>	<i>alag-ulu</i>
<i>two</i>	<i>child-DU</i>

'two kids.' (1998: 75)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL* (1998: 73-76)

6. Obligatoriness of number marking: *DU*: obligatory (1998: 73); *PL*: optional (1998: 76)

7. Non-numeral quantifiers: *Reduplication*, e.g. *bungmaji(SG)/bungmungmaji(PL)* 'old man' (1998: 43, 106)

Classifier systems

8. Noun classes: *Present (WALS)*

9. Numeral classifiers: *Absent (1998: passim)*

Language: Waorani
Family/Genus: Waorani/Waorani
Country/ Macro Area: Ecuador/South America
Reference(s): Aikhenvald, A. (2000) and Peeke, C. (1994)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:
{*N, NSG*}
tōdīya-da
sibling-3DU
'two brothers' (1994: 269)
4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL (1994: passim)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*
9. Numeral classifiers: *Obligatory (WALS)*

Language: Wappo
Family/Genus: Wappo-Yukian
Country/ Macro Area: United States (California)/North America
Reference(s): Li, C. N and Thomson, S.A. (1977) and Nichols, J. (1992)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:
(*N,NUM*)
chica hopóka
bear three
'three bear' (1977: 87)
4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (1997: 88)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (1992: 299)*
9. Numeral classifiers: *Absent (WALS)*

Language: Warao
Family/Genus: Warao/Warao
Country/ Macro Area: Venezuela/South America
Reference(s): Romero-Figeroa, A. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

nobo isaka

grandfather one

'one grandfather' (2003: 53)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (2003: 53, 98)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

bare kaika warao orabakaya ha

father(priest) with Warao four COP

'Four Warao are with the priest (they are under the priest's responsibility). (2003: 90)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (2003: 90, 82)*

Number systems

5. Number distinction: *SG/PL (2003: 43)*

6. Obligatoriness of number marking: *Obligatory (2003: 43)*

7. Non-numeral quantifiers: *LQ, e.g. reko 'both' as in noboto-ma reko [child-PL both] 'both children' (2003: 55)*

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Absent (2003: passim)*

Language: Warekena
Family/Genus: Arawakan/Arawakan
Country/ Macro Area: Brazil, Colombia, Venezuela/South America
Reference(s): Aikhenvald, A. (1998)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

ba-buya *pepuŋi*

one-CLF *day*

'one day' (1989: 299)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions* (1989: 265, 299)

3. Structural patterns of CNNC_{NSG}

{N,NUM}

te Jetfi *ibu*

three *head*

'three heads' (1989: 304)

{N,NUM,NSG}

kwatru *ne ŋima-pe*

four *cousin-PL*

'four cousins' (1989: 304)

4. Restrictions on the patterns in (3) above:

{N,NUM}: *Used with a non-personal reference* (1989: 304)

{N,NUM,NSG}: *Used with a personal reference (NB: PL optional)* (1989: 304)

Number systems

5. Number distinction: *SG/PL* (1989: 304)

6. Obligatoriness of number marking: *Optional* (1989: 304)

7. Non-numeral quantifiers: *LQ*, e.g. *ŋupe* 'many', as in *ŋupe kuluwitu* [*many vulture*] 'many vultures' (1989: 251)

Classifier systems

8. Noun classes: *Present* (1989: 299)

9. Numeral classifiers: *Present* (1989: 299)

Language: Warembori
Family/Genus: Lower Mamberamo/Lower Mamberamo
Country/ Macro Area: Indonesia/Australia-New Guinea
Reference(s): Donohue, M. (1999b) and personal communication*

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NSG}

ane-pa-kaindu

crocodile-big.ATTR-DU

'two big crocodiles' (1999b: 21)

{N,NUM}

mani-yave wonti

bird-DEF three

'(There were) three birds.' (1999b: 24)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL (1999b:10)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: *LQ, e.g. tekara 'many', as in anta-na tekara-na [fish-PL many-PL] 'lot's of fish' (1999: 62)*

Classifier systems

8. Noun classes: *Absent (Mark Donohue, p.c.)*
9. Numeral classifiers: *No information*

* Mark Donohue is Professor of Linguistics at the University of Monash, Australia. Email: Mark.Donohue@arts.monash.edu.au

Language: Wari'
Family/Genus: Chapacura-Wanhan/Chapacura-Wanhan
Country/ Macro Area: Brazil/South America
Reference(s): Everett, D. and Kern, B. (1997)

CNNCs

1. Structural patterns of CNNC_{SG}: *No CNNC_{SG} proper but a construction consisting of a noun and a word implying one-ness.*

<i>xica'</i>	<i>pe</i>	<i>na</i>	<i>tarama'</i>
<i>alone</i>	<i>be.at.SBJ</i>	<i>3SG.RP/P</i>	<i>man</i>

'There is one man.' (lit. *'The man is alone.'*) (1997: 348)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}: *No CNNC_{NSG} proper but a construction consisting of a noun and a word implying two-ness.*

<i>'tucu</i>	<i>caracan</i>	<i>na</i>	<i>xirim</i>
<i>face-1SBJ</i>	<i>each.other</i>	<i>3SG.RP/P</i>	<i>house</i>

'There are two houses.' (lit. *'The houses face each other.'*) (1997: 348)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction*

6. Obligatoriness of number marking: *See (5)*

7. Non-numeral quantifiers: *LQ, e.g. mijo 'many'(1997: 291); Collective marker, e.g. 'oro-pana (COLL-tree) 'trees' (1997: 290); collective nouns, e.g. hwijima 'children'*

Classifier systems

8. Noun classes: *Present (1997: 294)*

9. Numeral classifiers: *Absent (1997: 300)*

Language: Welsh
Family/Genus: Indo-European/Celtic
Country/ Macro Area: United Kingdom (Wales)/Eurasia
Reference(s): Thorne, D. (1993)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

un ferch

one girl

'one girl' (1993: 30)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,OBL,NSG}

naw o ddynion

nine of man.PL

'nine men' (1993: 149)

{N,NUM}

dwy ferch

two girls

'two girls' (1993: 31)

4. Restrictions on the patterns in (3) above:

{N,NUM,OBL,NSG}: *Used with high-valued numerals (1993: 149)*

{N,NUM}: *Used with low-valued numerals (1993: 149)*

Number systems

5. Number distinction: *SG/PL (1993: passim)*

6. Obligatoriness of number marking: *Obligatory (1993: passim)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (1993: passim)*

9. Numeral classifiers: *Absent (1993: passim)*

Language: Wichita
Family/Genus: Caddoan/Caddoan
Country/ Macro Area: United States (West Central Oklahoma) /North America
Reference(s): Rood, D. S. (1976)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>tá:kwicha</i>	<i>hasʔ</i>	<i>a:ʔa:ki</i>	<i>weʔes</i>	<i>ʔarhi</i>
<i>four</i>	<i>NARR</i>	<i>QUOT.AOR.3SG dog</i>	<i>be.a.number</i>	

'There were four dogs.' (1976: 13)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL (1976: 115)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*
9. Numeral classifiers: *No information*

Language: Wolof
Family/Genus: Niger-Congo/Nothern Atlantic
Country/ Macro Area: Gambia, Senegal/Africa
Reference(s): Ngom, F. (2003)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

<i>bénn</i>	<i>xarit</i>
<i>one</i>	<i>friend</i>

'one friend' (2003: 48)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>ñaar-i</i>	<i>xarit</i>
<i>two-PL</i>	<i>friend</i>

'two friends' (2003: 48)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2003: 19)*

6. Obligatoriness of number marking: *Obligatory (2003: 19)*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (2003: 19)*

9. Numeral classifiers: *Absent (2003: passim)*

Language: Yagua
Family/Genus: Peba-Yaguan/Peba-Yaguan
Country/Macro Area: Peru/South America
Reference(s): Nichols, J. (1992) and Payne, D. L. (1985)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

<i>tá-siy-quii</i>	<i>dáanta</i>
<i>one-CLF-one</i>	<i>medicine</i>
<i>'one pill' (1985: 172)</i>	

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,CLF}

<i>dá-nu-juɣ</i>	<i>ravichuɣ</i>
<i>two-CLF-two</i>	<i>rock</i>
<i>'two rocks' (1985: 184)</i>	

{N,NUM,CLF,NSG}

<i>dá-nuuy-juɣ</i>	<i>vánu-juɣ</i>	<i>rimityuvuɣ-juɣ</i>
<i>two-CLF-DU</i>	<i>man-DU</i>	<i>old.one-DU</i>
<i>'two old men' (1985: 112)</i>		

{N,NSG}

<i>núcovaañu-juɣ</i>
<i>wasp-DU</i>
<i>'two wasps' (1985: 320)</i>

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

- 5. Number distinction: *SG/DU/PL (1985: passim)*
- 6. Obligatoriness of number marking: *No information*
- 7. Non-numeral quantifiers: -

Classifier systems

- 8. Noun classes: *Present (1992: 301)*
- 9. Numeral classifiers: *Obligatory (WALS)*

Language: Yale (Kosarek)
Family/Genus: Trans-New Guinea/Mek
Country/ Macro Area: Indonesia/Australia-New Guinea
Reference(s): Heesch, V. (1992)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

nimi nhon

man one

'a/one man' (1992: 35)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

kel-abo pende

woman-PL two

'two women' (1992: 29)

{N,NUM}

nimi pende

man two

'two men' (1992: 30)

4. Restrictions on the patterns in (3) above:

{N,NUM,NSG}: *Used with human nouns and kinship terms (1992: 29)*

{N,NUM}: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (NB: only human nouns) (1992: 29)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: *LQ, e.g. meikno 'many', as in nimi meikno [man a.lot]'a lot of men' (1992: 20)*

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Yaqui
Family/Genus: Uto-Aztecan/Cahita
Country/ Macro Area: Mexico /North America
Reference(s): Dedrick, J. M. and E. H. Casad (1999) and Lindenfeld, J. (1973)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

wepul o'oo

one man

'one man' (1973: 28)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (1973: 28, 47)*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

naiki wakes-im

four cow-PL

'four cows' (1999: 233)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (1973: 49, 50)*

Number systems

5. Number distinction: *SG/PL (1999: 132)*

6. Obligatoriness of number marking: *Obligatory (WALS)*

7. Non-numeral quantifiers: *LQ, e.g. hwebena 'many', as in hwebena sewa-m [many flower-PL] 'a lot of flowers' (1973: 28)*

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *No information*

Language: Yelŋ Dnye
Family/Genus: Yele/Yele
Country/ Macro Area: Papua New Guinea/Australia-New Guinea
Reference(s): Henderson, J. (1995)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

<i>tpile</i>	<i>ngamê</i>
<i>thing</i>	<i>one</i>
<i>'one thing' (1995: 60)</i>	

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>póódó</i>	<i>miyó</i>
<i>rope</i>	<i>two</i>
<i>'two ropes' (1995: 59)</i>	

{N,NSG}

<i>tp:oo</i>	<i>dê</i>
<i>3SG.POSS.son</i>	<i>DU</i>
<i>'his two sons' (1995: 60)</i>	

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL (1995: passim)*

6. Obligatoriness of number marking: *No information*

7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *No information*

9. Numeral classifiers: *No information*

Language: Yidiny
Family/Genus: Australian/Pama-Nyungan
Country/ Macro Area: Australia (Queensland)/Australia-New Guinea
Reference(s): Dixon, R.M.W. (1977)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

ɲayu bama wawa:diɲu dambu:l

1SG.SA person.ABS see.PST two.ABS

I (unexpectedly) saw two people (1977: 270)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *No distinction (WALS)*
6. Obligatoriness of number marking: *See (5)*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*
9. Numeral classifiers: *Absent (WALS)*

Language: Yimas
Family/Genus: Lower Sepik-Ramu/Lower Sepik
Country/ Macro Area: Papua New Guinea/Australia-New Guinea
Reference(s): Foley, W. (1991)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,SG}

<i>tanm</i>	<i>mpa-m</i>
<i>bone.VII</i>	<i>one-VII</i>

'one bone' (1991: 101) (Class -m = singular number)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM,NSG}

<i>tanpl</i>	<i>p-rpal</i>
<i>bone.VII.DU</i>	<i>VII-two</i>

'two bones' (1991: 101)

<i>tanpat</i>	<i>p-ramnawt</i>
<i>bone.VII.PL</i>	<i>VII-three</i>

'three bones' (1991: 101)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

- 5. Number distinction: *SG/DU/PL (1991: 91, 120)*
- 6. Obligatoriness of number marking: *Obligatory (1991: 2)*
- 7. Non-numeral quantifiers: -

Classifier systems

- 8. Noun classes: *Present (1991: 2)*
- 9. Numeral classifiers: *Absent (1991: 21)*

Language: Yuchi
Family/Genus: Yuchi/Yuchi
Country/ Macro Area: United States (Tennessee)/North America
Reference(s): Linn, M. S. (2002)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM}

<i>k'as'æthechya</i>	<i>nōwē</i>
<i>car</i>	<i>two</i>
<i>'two cars.' (2002: 443)</i>	

{N,NUM,NSG}

<i>got'ē</i>	<i>nōwē</i>	<i>hōñō</i>
<i>man</i>	<i>two</i>	<i>PL</i>
<i>'two men' (2002: 479)</i>		

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/PL (2002: 379)*
6. Obligatoriness of number marking: *No information*
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Present (2002: passim)*
9. Numeral classifiers: *Absent (WALS)*

Language: Yukaghir (Kolyma)
Family/Genus: Yukaghir/Yukaghir
Country/ Macro Area: Russia (Siberia)/Eurasia
Reference(s): Maslova, E. (2003a)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

irk-in *aŋd'e*
one-ATTR *eye*
 'one eye' (2003: 82)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

ataq-un *tuis*
two-ATTR *basket*
 'two baskets' (2003: 83)

{N,NUM,NSG}

purk-in *čül'd'ŋ* *pulut-pe-lek* *kel-ŋi-l*
seven-ATTR *fairy.tale* *old.man-PL-PRED* *come-3PL-SF*
 '...Seven ogres came.' (2003: 85)

4. Restrictions on the patterns in (3) above:

{N,NUM}: *No known restrictions*
 {N,NUM,NSG}: *Rare (2003: 85), hence @*

Number systems

- 5. Number distinction: *SG/PL (2003: 3)*
- 6. Obligatoriness of number marking: *No information*
- 7. Non-numeral quantifiers: *LQ, e.g. čumut 'all', as in køj-pe čumut [man-PL all] 'all men' (2003: 75)*

Classifier systems

- 8. Noun classes: *Absent (WALS)*
- 9. Numeral classifiers: *Absent (WALS)*

Language: Yukaghir (Tundra)
Family/Genus: Yukaghir/Yukaghir
Country/ Macro Area: Russia (Siberia)/Eurasia
Reference(s): Maslova, E. (2003b)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM}

māraq-d *uo*
one-ATTR *child*
'one child' (2003: 45)

2. Restrictions on the patterns in (1) above: *No known restrictions*

3. Structural patterns of CNNC_{NSG}:

{N,NUM}

taŋ *jā-n* *gode-k* *ta-dā* *l'e-ŋu-l*
DST *three-ATTR* *person-FOC* *DAT-ADV* *be-PL-SF*
'There were only those three people.' (2003: 46)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

- 5. Number distinction: *SG/PL (2003: passim)*
- 6. Obligatoriness of number marking: *No information*
- 7. Non-numeral quantifiers: -

Classifier systems

- 8. Noun classes: *Absent (WALS)*
- 9. Numeral classifiers: *Absent (WALS)*

Language: Yup'ik
Family/Genus: Eskimo-Aleut/ Eskimo-Aleut
Country/ Macro Area: United States (Alaska)/North America
Reference(s): Mather, E., Marie, M. and Miyaoka, O. (2002)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N.NSG}

<i>arna-k</i>	<i>iqvartuk</i>
<i>woman-ABS.DU</i>	<i>pick.berry.IND.3DU</i>

'The two women are picking berries.' (2002: 58)

{N,NUM,NSG}

<i>malru-k</i>	<i>kipusvii-k</i>
<i>two-ABS.DU</i>	<i>store-ABS.DU</i>

'two stores' (2002: 73)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL* (2002: 55)
6. Obligatoriness of number marking: *Obligatory* (2002: 56)
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent (WALS)*
9. Numeral classifiers: *No information*

Language: Yurok
Family/Genus: Algic/Yurok
Country/ Macro Area: United States (California)/North America
Reference(s): Robins, R. H. (1958)

CNNCs

1. Structural patterns of CNNC_{SG}:

{N,NUM,CLF}

kohtoh *hegor*

one.CLF *month*

'one month' (1958: 89)

2. Restrictions on the patterns in (1) above: *Used generally with no known restrictions (1958: 87)*

3. Structural patterns of CNNC_{NSG}

{N,NUM,CLF}

ni?iʔ *peg.ʔk*

two.CLF *man*

'two men' (1958: 86)

4. Restrictions on the patterns in (3) above: *Used generally with no known restrictions (1958: 87)*

Number systems

5. Number distinction: *SG/PL (1958: 23)*

6. Obligatoriness of number marking: *Optional (1958: 23)*

7. Non-numeral quantifiers: *Reduplication, e.g. slek^woh (SG)/slek^wslek^w(PL) 'cloth(es)' (1958: 13-14)*

Classifier systems

8. Noun classes: *Absent (WALS)*

9. Numeral classifiers: *Obligatory (1958: 86-87)*

Language: Zuni
Family/Genus: Zuni/Zuni
Country/ Macro Area: United States (New Mexico)/North America
Reference(s): Nichols, L. (1997/2003)

CNNCs

1. Structural patterns of CNNC_{SG}: *No information*
2. Restrictions on the patterns in (1) above: *No information*
3. Structural patterns of CNNC_{NSG}:

{N,NUM,SG}

<i>ho'</i>	<i>ha'i</i>	<i>'e'ni-nne</i>	<i>'illi</i>
<i>1SG.NOM</i>	<i>three</i>	<i>belt-SG</i>	<i>have</i>

'I have three belts.' (1997: 12)

4. Restrictions on the patterns in (3) above: *No known restrictions*

Number systems

5. Number distinction: *SG/DU/PL* (1997: 12)
6. Obligatoriness of number marking: *Obligatory* (WALS)
7. Non-numeral quantifiers: -

Classifier systems

8. Noun classes: *Absent* (WALS)
9. Numeral classifiers: *Absent* (WALS)

References

- Abondolo, D. M. (1998). *The Uralic languages*. London: Routledge.
- Acquaviva, P. (2006). Goidelic inherent plurals and the morphosemantics of number. *Lingua*, 116 (11), 1860-1887.
- Adams, K. L. (1991). The influence of non-Austroasiatic languages on numeral classification in Austroasiatic. *Journal of the American Oriental Society*, 111(1), 62-81.
- Adelaar, W. F. H., & Muysken, P. (2004). *The languages of the Andes*. Cambridge: Cambridge University Press.
- Aikhenvald, A. (1998). Warekena. In D. C. Derbyshire & G. K. Pullum (Eds.), *Handbook of Amazonian languages* (Vol. 4, pp. 215-439). Berlin: Mouton de Gruyter.
- Aikhenvald, A. (2000). *Classifiers: a typology of noun categorization devices*. Oxford: Oxford University Press.
- Aikhenvald, A. (2002). *Language contact in Amazonia*. Oxford: Oxford University Press.
- Aikhenvald, A. (2003). *A grammar of Tariana*. Cambridge: Cambridge University Press.
- Aikhenvald, A. (2006). Classifiers and noun classes: semantics. In A. Anderson & E. K. Brown (Eds.), (pp. 463-471).
- Aikhenvald, A. & Dixon, R. M. W. (Eds.). (2006). *Areal diffusion and genetic inheritance: problems in comparative linguistics*. Oxford: Oxford University Press.
- Allan, K. (1977). Classifiers. *Language*, 53, 284-310.
- Ambrasas, V. (1997). *Lithuanian grammar*. Vilnius: Baltos Lankos.
- Amha, A. (2001). *The Maale language*. Leiden: CNWS
- Anderson, A. & Brown, E. K. (2006). *Encyclopedia of language & linguistics*. 2nd. from <http://ezproxy.lib.ed.ac.uk:2048/login?url=http://www.sciencedirect.com/science/referenceworks/0080448542>
- Anderson, A. et al. (2006). São Tomé and Príncipe: language situation. In A. Anderson & E. K. Brown (Eds.), (pp. 751-752).
- Anderson, G.D.S. (2006). Austroasiatic languages. In A. Anderson & E. K. Brown (Eds.), (pp. 598-600).
- Anderson, J. L. (1989). *Comaltepec Chinantec syntax*. Dallas, TX [Arlington, Tex.]: Summer Institute of Linguistics; University of Texas at Arlington.
- Andrade, M. J. (1933). *Quileute*. New York: Columbia University Press.
- Antunes de Araujo, G. (2004). *A grammar of Sabanê: a Nambikwaran language*. Utrecht: LOT.
- Asher, R. E., & Simpson, J. M. Y. (1994). *The encyclopedia of language and linguistics*. New York: Pergamon Press.
- Austin, P. (1981). *A grammar of Diyari, South Australia*. Cambridge: Cambridge University Press.

- Bailey, B. L. (1966). *Jamaican Creole syntax: a transformational approach*. Cambridge: Cambridge University Press.
- Bailey, C. J. N. (1982). *On the Yin and Yang nature of language*. Ann Arbor: Karoma.
- Bak, P. & Paczuski, M. (1996). Mass extinctions vs. uniformitarianism in biological evolution [Electronic Version]. Retrieved 1 December 2007 from http://arxiv.org/PS_cache/cond-mat/pdf/9602/9602012v1.pdf.
- Bakker, P. & Parkvall, M. (2002). Reduplication in pidgins and creoles [Electronic Version]. Retrieved 25 June 2005 from <http://www-gewi.kfunigraz.ac.at/ling/veranst/redup2002/abstracts/bakker-parkwall.pdf>.
- Bakro-Nagy, M. (2005). The Uralic language family: facts, myths and statistics (Review). *Lingua*, 115(7), 1053-1062.
- Ball, M. J. & Fife, J. (1993/2002). *The Celtic languages*. London: Routledge.
- Ball, M. J., & Mèuller, N. (1992). *Mutation in Welsh*. London: Routledge.
- Barnes, J. (1999). Tucano. In R. M. W. Dixon & A. Aikhenvald (Eds.), *The Amazonian languages* (pp. 207-226). Cambridge: Cambridge University Press.
- Bateman, J. (1986). *Iau verb morphology*. Jakarta: Badan Penyelenggara Seri Nusa, Universitas Katolik Indonesia Atma Jaya.
- Bauer, W., Parker, W., & Evans, T. K. (1993). *Maori*. London: Routledge.
- Bee, D. (1973). Usarufa: a descriptive grammar. In H. McKaughan (Ed.), *The languages of the eastern family of the east New Guinea highland stock* (pp. 324-400). Seattle: University of Washington Press.
- Bell, C. R. V. (1953). *The Somali language*. London: Longmans, Green.
- Beller, R., & Beller, P. (1979). Huasteca Nahuatl. In R. W. Langacker (Ed.), *Modern Aztec grammatical sketches* (pp. 199-306). Dallas: Summer Institute of Linguistics.
- Benmamoun, E. (2003). Agreement parallelism between sentences and noun phrases: a historical sketch. *Lingua*, 113(8), 747-764.
- Bentley, R. A., Hahn, M. W., & Shennan, S. J. (2004). Random drift and culture change. *Proceedings of the royal society, B* 271, 1443- 1450.
- Berg, H. (1995). *A grammar of Hunzib (with texts and lexicon)*. München: Lincom Europa.
- Berlin, B., & Kay, P. (1969). *Basic color terms: their universality and evolution*. Berkeley: University of California Press.
- Berry, K (1995). *A description of the Abun language: phonology and basic morpho-syntax*. Unpublished Thesis (M.A.) - La Trobe University.
- Besnier, N. (2000). *Tuvaluan: a Polynesian language of the central pacific*. London: Routledge.
- Bibiko, H.-J. (2005). *Interactive Reference Tool of the World Atlas of Language Structures (WALS)*. Oxford: Oxford University Press.
- Bickerton, D. (1981). *Roots of language*. Ann Arbor: Karoma.
- Bisang, W. (1999). Classifiers in East and Southeast Asian languages: counting and beyond. In J. Gvozdanović (Ed.), (pp. 113-185).
- Blackings, M. J., & Fabb, N. (2003). *A grammar of Ma'di*. Berlin: Mouton De Gruyter.
- Blake, B. J. (1994). *Case*. Cambridge: Cambridge University Press.
- Blanc, H. (1970). Dual and pseudo-dual in the Arabic dialects. *Language*, 46(1), 42-57.

- Blevins, J. (2004). The morphology of Yurok numerals [Electronic Version], abstract. Retrieved 6 June 2006 from <http://monolith.eva.mpg.de/~gil/numerals/abstracts/Blevins.pdf>
- Bonnerjea, R. (1978). A comparison between Eskimo-Aleut and Uralo-Altaic demonstrative elements, numerals, and other related semantic problems. *International journal of American linguistics*, 44(1), 40-55.
- Bonvillain, N. (1973). *A grammar of Akwesasne Mohawk*. Ottawa: National Museums of Canada.
- Borer, H. (2005). *Structuring sense*. Oxford: Oxford University Press.
- Borg, A. J. & Azzopardi-Alexander, M. (1997). *Maltese*. London: Routledge.
- Borgman, D. M. (1990). Sanuma. In D. C. Derbyshire & G. K. Pullum (Eds.), *Handbook of Amazonian languages* (Vol. 2, pp. 15-248). Berlin: Mouton de Gruyter.
- Bowden, J. (2001). *Taba: description of a South Halmahera language*. Canberra: Pacific Linguistics.
- Braine, J. C. (1970/2000). *Nicobarese grammar (Car dialect)*. Ann Arbor: UMI.
- Briley, D. (1997). Four grammatical marking systems in Bauzi. In K. Franklin (Ed.), *Pacific Linguistics Series A-85 (Papers in Papuan Linguistics No. 2)*. Canberra: Australian National University.
- Brinton, L. J. & Traugott, E. C. (2005). *Lexicalization and language change*. Cambridge: Cambridge University Press.
- Brown, M. C. (2004). *Topics in Iquito syntax: word order, possession, and nominal discontinuity*. Unpublished Thesis (M.A.)- University of Texas, Austin.
- Burrow, T. (2001). *The Sanskrit language*. Retrieved 16 April 2007, from <http://books.google.co.uk/books?id=cWDhKTj1SBYC&printsec=frontcover&dq=burrow+sanskrit&sig=B4EyKdyn30u7gfDyCZnvdj9u1vg>.
- Bybee, J. L. (2003). Mechanisms of change in grammaticalization: the role of frequency. In B. D. Joseph & R. D. Janda (Eds.), (pp. 602-623).
- Bybee, J. L., Perkins, R. D. & Pagliuca, W. (1994). *The evolution of grammar: tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- Caldwell, R. et al. (2006). *Understanding evolution*. Retrieved August 2007, from www.ucmp.berkeley.edu.
- Camaj, M. (1984). *Albanian grammar, with exercises, chrestomathy, and glossaries*. Wiesbaden: O. Harrassowitz.
- Campbell, G. L. (Ed.). (2000). *Compendium of the world's languages* (2nd ed. Vol. 1 and 2). London: Routledge.
- Capell, A., & Hinch, H. E. (1970). *Maung grammar: texts and vocabulary*. The Hague: Mouton.
- Cardona, G., & Jain, D. (2003). *The Indo-Aryan languages*. London: Routledge.
- Casper, F. B., & Schäfer, F. S. (1944). *A grammar of the Miskito language*. Marathon, Wis.: The Divina Pastora Society.
- Centre for Information on Language Teaching and Research. (1986). *Norwegian*. London: Centre for Information on Language Teaching and Research.
- Charlesworth, D., & Charlesworth, B. (2003). *Evolution: a very short introduction*. New York: Oxford University Press.

- Chatterji, S. K. (1926). *The origin and development of the Bengali language*. Calcutta: Calcutta University Press.
- Cheshire, J. (1982). *Variations in an English dialect: a sociolinguistic study*. Cambridge: Cambridge University Press.
- Chirkova, K. (2004). One item to ten items in written and spoken Mandarin [Electronic Version], abstract. Retrieved 6 June 2006 from <http://monolith.eva.mpg.de/~gil/numerals/abstracts/Chirkova.pdf>.
- Chung, C.-H., & Chung, K.-J. (1996). Kuot grammar essentials. In J. M. Clifton (Ed.), *Two Non-Austronesian grammars from the islands*. Ukarumpa, EHP, Papua New Guinea: SIL
- Clackson, J. P. T. (2004). Latin. In R. D. Woodard (Ed.), *The Cambridge encyclopedia of the world's ancient languages*. Cambridge: Cambridge University Press.
- Clifton, J. M. (1997). The Kaki Ae language. In S. A. Wurm (Ed.), (pp. 3-65): Canberra: Australian National University.
- Closs, M. P. (1986). *Native American mathematics*. Austin: University of Texas Press.
- Coelho, G. M. (2003). *A grammar of Betta Kurumba*. Austin: University of Texas Press.
- Comrie, B. (1987). *The world's major languages*. London: Croom Helm.
- Comrie, B. (1999). Haruai numerals and their implications for the history and typology of numeral systems. In J. Gvozdanović (Ed.), (pp. 81-94).
- Comrie, B., Haspelmath, M., & Bickel, B. (2008). *The Leipzig glossing rules: conventions for interlinear morpheme-by-morpheme glosses* [Electronic Version], 1-10 from <http://www.eva.mpg.de/lingua/pdf/LGR08.02.05.pdf>.
- Comrie, B., & Kuteva, H. (2005). The evolution of grammatical structures and “functional need” explanations. In M. Tallerman (Ed.), *Language origins: perspectives on evolution*, (pp. 185–207). Oxford: Oxford University Press.
- Comrie, B., & Smith, N. (1977). Lingua descriptive studies: questionnaire. *Lingua*, 42(1), 1-71.
- Conant, L. L. (1896). *The number concept: its origin and development*. New York, London: Macmillan.
- Conrad, R. J., & Wogiga, K. (1991). *An outline of Bukiyip grammar*. Canberra: Australian National University.
- Conzemius, E. (1929). Notes on the Miskito and Samu languages of Eastern Nicaragua and Honduras. *International journal of American linguistics*(5), 57-115.
- Cook, E.-D. (1984). *A Sarcee grammar*. Vancouver: University of British Columbia Press.
- Cook, E.-D. (2004). *A grammar of Dëne (Chipewyan)*. Winnipeg: Algonquian and Iroquoian Linguistics.
- Corbett, G. G. (1978). Universals in the syntax of cardinal numerals. *Lingua*, 46(1), 61-74.
- Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press.
- Corbett, G. G. (2000). *Number*. Cambridge: Cambridge University Press.
- Corbett, G. G. (2005). Number of genders. In M. Haspelmath, M. Dryer, D. Gil & B. Comrie (Eds.), (pp. 126-129).
- Corver, N., Doetjes, J., & Zwarts, J. (2007). Linguistic perspectives on numerical expressions: introduction. *Lingua*, 117(5), 751-757.

- Cowan, H. M. (1965). *Grammar of the Sentani language (with specimen texts and vocabulary)*. Gravenhage: Martinus Nijhoff.
- Craig, C. (Ed.). (1986a). *Noun classes and categorization: proceedings of a symposium on categorization and noun classification, Eugene, Oregon, October 1983*. Amsterdam: Benjamins.
- Craig, C. G. (1986b). Jacalteco noun classifiers. *Lingua*, 70(4), 241-284.
- Craig, C. G. (1994). Classifier languages. In R. E. Asher & J. M. Y. Simpson (Eds.), (pp. 565-569).
- Croft, W. (2000). *Explaining language change: an evolutionary approach*. Harlow: Longman.
- Croft, W. (2003). *Typology and universals* (2nd ed.). Cambridge: Cambridge University Press.
- Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Crothers, J. (1978). Typology and universals of vowel systems. In J. H. Greenberg, C. A. Ferguson & E. A. Moravcsik (Eds.), *Universals of human language: Phonology* (Vol. 2, pp. 93-52). Stanford: Stanford University Press.
- Crowley, T. (1998). *An Erromangan (Sye) grammar*. Honolulu: University of Hawaii Press.
- Crowley, T. (2004). *Bislama reference grammar*. Honolulu: University of Hawai'i Press.
- Crum, B., & Dayley, J. P. (1993). *Western Shoshoni grammar*. Boise: Boise State University.
- Cruse, D. A. (1994). Number and number systems. In R. E. Asher & J. M. Y. Simpson (Eds.), (pp. 2857-2861).
- Curnow, T. J. (1997). *A grammar of Awa Pit (Cuaiquer): an indigenous language of south-western Colombia*. Unpublished Thesis (Ph.D)-Australian National University.
- Dahl, Ö. (2004). *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.
- Dahl, Ö., & Koptjevskaja-Tamm, M. (2001). *The Circum-Baltic languages: typology and contact*. Amsterdam: John Benjamins.
- Darwin, C., & Burrow, J. W. (1968). *The origin of species by means of natural selection or the preservation of favoured races in the struggle for life*. Harmondsworth: Penguin.
- Dasgupta, P., & Jain, D. (2003). Bangla. In G. Cardona & D. Jain (Eds.), (pp. 351-390).
- Davidson, M. (2002). *Studies in southern Wakashan (Nootkan) grammar*. Unpublished Thesis (Ph.D.)- Buffalo University.
- Davies, W. (1999). *Madurese*. München: LINCOM
- Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.
- Day, S. B., & Gentner, D. (2007). Nonintentional analogical inference in text comprehension. *Memory & Cognition*, 35(1), 39-49.
- de Boer, B. (2001). *The origins of vowel systems*. Oxford: Oxford University Press.
- de Vries, J. A. (1997). An overview of Kwerba verb morphology. In A. Pawley (Ed.), *Papers in Papuan Linguistics No. 3* (pp. 1 – 35).

- de Vries, L. (1993). *Forms and functions in Kombai, an Awyu language of Irian Jaya*. Canberra: Australian National University.
- de Vries, L. (2004). *A short grammar of Inanwatan, an endangered language of the Bird's Head of Papua, Indonesia*. Canberra: Pacific Linguistics.
- Dedrick, J. M., & Casad, E. H. (1999). *Sonora Yaqui language structures*. Tucson: University of Arizona Press.
- Dehaene, S. (1997). *The number sense: how the mind creates mathematics*. Oxford: Oxford University Press.
- Denwood, P. (1999). *Tibetan*. Amsterdam: John Benjamins.
- Dik, S. C., & Hengeveld, K. (1997). *The theory of functional grammar (Part 1, Structure of the clause)* (2nd rev. ed.). Berlin: Mouton de Gruyter.
- Dixon, R. M. W. (1972). *The Dyirbal language of North Queensland*. London: Cambridge University Press.
- Dixon, R. M. W. (1977). *A grammar of Yidiñ*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. (1986). Noun classes and noun classification in typological perspective. In C. Craig (Ed.), (pp. 105-112).
- Dixon, R. M. W. (1997). *The rise and fall of languages*. Cambridge: Cambridge University Press.
- Dixon, R. M. W. (2002). *Australian languages: their nature and development*. Cambridge: Cambridge University Press.
- Dixon, R. M. W., & Vogel, A. R. (2004). *The Jarawara language of Southern Amazonia*. Oxford: Oxford University Press.
- Dobson, W. A. C. H. (1962). *Early Archaic Chinese: a descriptive grammar*. Toronto: University of Toronto Press.
- Donaldson, T. (1980). *Ngiyambaa: the language of the Wangaaybuwan*. Cambridge: Cambridge University Press.
- Donohue, M. (1999a). *A grammar of Tukang Besi*. Berlin: Walter De Gruyter
- Donohue, M. (1999b). *Warembori*. München: Lincom Europa.
- Downing, P. (1996). *Numeral classifier systems: the case of Japanese*. Amsterdam-Philadelphia: John Benjamins.
- Doyle, A. (2001). *Irish*. München: Lincom Europa.
- Dryer, M. S. (1989a). Large linguistic areas and language sampling. *Studies in Language* (13), 257-292.
- Dryer, M. S. (1989b). Plural Words. *Linguistics* (27), 865-895.
- Dryer, M. S. (1992). The Greenbergian word order correlations. *Language* 68, 81-138.
- Dryer, M. S. (2003). Significant and non-significant implicational universals. *Linguistic typology* 7, 108-128.
- Dryer, M. S. (2005a). Coding of nominal plurality. In M. Haspelmath, M. Dryer, D. Gil & B. Comrie (Eds.), (pp. 138-141).
- Dryer, M. S. (2005b). Order of Numeral and Noun. In M. Haspelmath, M. Dryer, D. Gil & B. Comrie (Eds.), (pp. 362-365).

- Dunbar, K. (1999). How scientists build models in Vivo science as a window on the scientific mind. In L. Magnani, N. J. Nersessian & P. Thagard (Eds.), *Model-based reasoning in scientific discovery* (pp. 85-100). New York: Kluwer Academic/Plenum Publishers.
- Dunn, J. A. (1979). *A reference grammar for the Coast Tsimshian language*. Ottawa: National Museums of Canada.
- Dutton, T. E. & Thomas, D. (1985). *A new course in Tok Pisin (New Guinea Pidgin)*. Canberra: Australian National University.
- Ebert, K. H. (2003). Camling. In G. Thurgood & R. J. LaPolla (Eds.), *The Sino-Tibetan languages* (pp. 533-545). London: Routledge.
- Elders, S. (2000). *Grammaire Mundang*. Leiden: Research School of Asian, African, and Amerindian Studies, Universiteit Leiden.
- Emeneau, M. B. (1956). India as a Linguistic Area. *Language*, 32(1), 3-16.
- Engelenhoven, A. T. (2004). *Leti: a language of Southwest Maluku*. Leiden: KITLV Press.
- England, N. C. (1983). *Grammar of Mam: A Mayan language*. Austin: Texas University Press.
- Epps, P. (2004). Tracing the development of numerals in the Guaviaré-Japurá (Maku) family [Electronic Version], abstract. Retrieved 24 June 2006 from <http://monolith.eva.mpg.de/~gil/numerals/abstracts/Epps.pdf>
- Evans, N. (1995). *A grammar of Kayardild: with historical-comparative notes on Tangkic*. Berlin: Mouton de Gruyter.
- Everett, D. (2005). Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology*, 46(4), 621-646.
- Everett, D. L., & Kern, B. (1997). *Wari': the Pacaas Novos language of western Brazil*. London: Routledge.
- Fabian, G., Fabian, E., & Waters, B. (1998). *Morphology, syntax and cohesion in Nabak, Papua New Guinea* (Vol. 144). Canberra: Pacific Linguistics.
- Faraclas, N. (1996). *Nigerian Pidgin*. London: Routledge.
- Feldman, H. (1986). *A grammar of Awtuw*. Canberra: Australian National University.
- Fleck, D. W. (2004). *A grammar of Matsés*. Ann Arbor: UMI.
- Foley, W. A. (1991). *The Yimas language of New Guinea*. Stanford: Stanford University Press.
- Fox, A. (1995). *Linguistic reconstruction: an introduction to theory and method*. Oxford: Oxford University Press.
- Frajzyngier, Z. (2001). *A grammar of Lele*. Stanford: CSLI.
- Frank, P. (1990). *Ika syntax*. Dallas: SIL.
- Gabas, N. (2002). *A grammar of Karo, Tupi (Brazil)*. Ann Arbor: UMI.
- Gentner, D. (1983). Structure-mapping: a theoretical framework for analogy. *Cognitive science*, 7, 155-170.
- Gentner, D. (1999). Analogy. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 17-19). Cambridge, MA: MIT Press.
- Gentner, D., & Kurtz, K. J. (2006). Relations, objects, and the composition of analogies. *Cognitive science*, 30, 1-34.

- Gil, D. (2005). Numeral classifiers. In M. Haspelmath, M. Dryer, D. Gil & B. Comrie (Eds.), (pp. 226-229).
- Gil, D. (2006). Primitive languages. In A. Anderson, E. & K. Brown (Eds.), (pp. 91-93).
- Gillies, W. (1993). The Celtic languages. In J. Fife & M. J. Ball (Eds.) (pp. 145-227).
- Givón, T. (1976). Topic, pronoun, and grammatical agreement. In C. N. Li (Ed.), *Subject and topic* (pp. 149-188). New York: Academic Press.
- Givón, T. (1979). *On understanding grammar*. London: Academic Press.
- Givón, T. (1995). Isomorphism in the grammatical code: cognitive and biological considerations. In R. Simone (Ed.), *Iconicity in language* (pp. 47-76). Amsterdam: John Benjamins.
- Gorelova, L. M. (2002). *Manchu grammar*. Boston: Brill.
- Goudswaard, N. E. (2005). *The Begak (Iida'an) language of Sabah*. Unpublished Thesis (Ph.D.) - Vrije Universiteit, Amsterdam.
- Gould, S. J., & Vrba, E. S. (1982). Exaptation: a missing term in the science of form. *Paleobiology*, 8, 4-15.
- Granberry, J. (2004). *Modern Chitimacha (Sitimaxa)*. München: Lincom.
- Gravelle, G. (2002). Morphosyntactic properties of Meyah word classes. In G. P. Reesink (Ed.), *Languages of the eastern Bird's Head, Pacific linguistics 524* (pp. 109-180). Canberra: Australian National University.
- Green, J. (2006). Noun Classification in Dakelh [Electronic Version]. Retrieved 3 January 2007 from <http://www.ydli.org/dakinfo/nounclas.htm>.
- Greenberg, J. H. (Ed.). (1963). *Universals of language*. Cambridge, MA: MIT Press.
- Greenberg, J. H. (1972/1990). Numeral classifiers and substantive number: problems in the genesis of a linguistic type. In J. H. Greenberg, K. M. Denning & S. Kemmer (Eds.), *On language: selected writings of Joseph H. Greenberg* (pp. 166-193). Stanford: Stanford University Press.
- Greenberg, J. H. (1978). Generalizations about numeral systems. In J. H. Greenberg, C. A. Ferguson & E. A. Moravcsik (Eds.), *Universals of human language: word structure* (Vol. 3, pp. 250-295). Stanford: Stanford University Press.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41-58). New York: Academic Press.
- Grierson, G. A. (1904). *Linguistic survey of India: Mon-Khmer and Siamese-Chinese families (including Khassi and Tai)* (Vol. 2). Calcutta: Office of the Superintendent of Government Printing, India.
- Grimes, B. F. (2000). *Ethnologue* (14th ed.). Dallas: SIL.
- Grimes, C. E. (1991). *The Buru language of Eastern Indonesia*. Unpublished Thesis (Ph.D.)-Australian National University.
- Grinevald, C. (2002a). Making sense of nominal classification systems: noun classifiers and the grammaticalization. In I. Wischer & G. Diewald (Eds.), *New reflections on grammaticalization*, (Vol. 49, pp. 259-275). Amsterdam: John Benjamins.
- Grinevald, C. (2002b). Nominal classification in Movima. In *Current studies on South American languages: selected papers from the 50th International Congress of Americanists in Warsaw and the Spinoza Workshop on Amerindian Languages in Leiden, 2000* Leiden: Research School of Asian, African, and Amerindian Studies.

- Grondona, V. M. (1998/2004). *A grammar of Mocoví*. Ann Arbor: UMI
- Gruzdeva, E. (1998). *Nivkh*. München: Lincom Europa.
- Guillaume, A. (2004). *A grammar of Cavineña, an Amazonian language of Northern Bolivia*. Unpublished Thesis (Ph.D.) - La Trobe University.
- Guirardello-Damian, R. (1999). *A reference grammar of Trumai*. Texas: Rice University.
- Gusain, L. (2003). *Mewati*. München: Lincom Europa.
- Gvozdanović, J. (Ed.). (1999). *Numeral types and changes worldwide*. Berlin: Mouton de Gruyter.
- Haarmann, H. (1990). Language as a seismograph as acculturation: Sociolinguistic parameters as a language contact in the Asian context [Electronic Version]. *Journal of Asian Pacific communication*, 1, 71-86. Retrieved 20 July 2007 from http://www.google.co.uk/books?pg=PA85&lpg=PA82&dq=numeral+classifier+Japanese+borrowed&sig=QCSDhZDIFvA5w8Ajh6_W8XrJIIM&id=di7684-HkEkC&ots=7KbRHofjWJ#PPA84,M1.
- Haiman, J. (1985). *Iconicity in syntax: proceedings of a symposium on iconicity in syntax, Stanford, June 24-26, 1983*. Amsterdam: Benjamins.
- Hansson, I.-L. (2003). Akha. In G. Thurgood & R. J. LaPolla (Eds.), (pp. 236-251).
- Hardman, M. J. (1966). *Jaqaru: outline of phonological and morphological structure*. The Hague: Mouton.
- Hardman, M. J. (2000). *Jaqaru*. München: Lincom Europa.
- Harris, A. C. (1981). *Georgian syntax: a study in relational grammar*. Cambridge: Cambridge University Press.
- Harrison, S. P. (2003). On the limits of the comparative method. In B. D. Joseph & R. D. Janda (Eds.), (pp. 213-243).
- Harvey, M. (2001). *A grammar of Limilngan: a language of the Mary River region, Northern Territory, Australia*. Canberra: Australian National University.
- Harvey, M. (2002). *A grammar of Gaagudju*. Berlin: Mouton de Gruyter.
- Haspelmath, M. (1993). *A grammar of Lezgian*. Berlin: Mouton de Gruyter.
- Haspelmath, M. (1999). Optimality and diachronic adaptation. [Electronic Version]. *Zeitschrift für Sprachwissenschaft* Retrieved December 2006 from http://amor.cms.hu-berlin.de/~h0998dgh/ot/dat/haspelmath_DiaAdapt.pdf
- Haspelmath, M. (To appear). Frequency vs. iconicity in explaining grammatical asymmetries [Electronic Version]. *Cognitive linguistics*, 19. Retrieved 25 June 2007 from <http://email.eva.mpg.de/~haspelmt/Iconicity.pdf>
- Haspelmath, M. (2005). Occurrence of nominal plurality. In M. Haspelmath, M. Dryer, D. Gil & B. Comrie (Eds.), (pp. 142-145).
- Haspelmath, M., Dryer, M., Gil, D., & Comrie, B. (Eds.). (2005). *The world atlas of language structures*. Oxford: Oxford University Press.
- Hawkins, R., & Towell, R. (2001). *French grammar and usage* (2nd ed.). London: Arnold.
- Heath, J. (1999). *A grammar of Koyraboro (Koroboro) Senni: the Songhay of Gao, Mali*. Köln: Köppe.
- Heeschen, V. (1992). *A dictionary of the Yale (Kosarek) language (with sketch of grammar and English index)* Berlin: Dietrich Reimer.

- Heine, B. (1999). *The Ani: grammatical notes and texts*. Köln Institut für Afrikanistik, Univ. zu Köln
- Heine, B., & Kuteva, T. (2002a). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Heine, B., & Kuteva, T. (2002b). On the evolution of grammatical forms. In A. Wray (Ed.), *The transition to language* (pp. 376-397). Oxford: Oxford University Press.
- Heine, B., & Kuteva, T. (2007). *The genesis of grammar: a reconstruction*. Oxford: Oxford University Press.
- Heine, B., & Reh, M. (1984). *Grammaticalization and reanalysis in African languages*. Hamburg: H. Buske.
- Helinski, E. (1998). Selkup. In D. M. Abondolo (Ed.), (pp. 548-579).
- Henderson, J. (1995). *Phonology and grammar of Yele, Papua New Guinea*. Canberra, Australia: Australian National University.
- Hercus, L. A. (1994). *A grammar of the Arabana-Wangkangurru language, Lake Eyre Basin, South Australia*. Canberra: Australian National University.
- Hewitt, B. G. (1995). *Georgian: a structural reference grammar*. Amsterdam; Philadelphia: John Benjamins.
- Hewitt, B. G. (1996). *Georgian: a learner's grammar*. London: Routledge.
- Hewitt, B. G., & Khiba, Z. K. (1979). *Abkhaz*. Amsterdam: North-Holland.
- Hofling, C. A. (2000). *Itzaj Maya grammar*. Salt Lake City: University of Utah Press.
- Holes, C. (1995). *Modern Arabic: structures, functions, and varieties*. London: Longman.
- Holm, J. A. (1988). *Pidgins and creoles* (Vol. 2). Cambridge: Cambridge University Press.
- Holm, J. A. (2000). *An introduction to pidgins and creoles*. Cambridge: Cambridge University Press.
- Holmer, N. M. (1946). Outline of Cuna grammar. *International Journal of American Linguistics*, 12(4), 185-197.
- Holmes, B. (2007). Language 'mutations' affect least-used words [Electronic Version]. *NewScientist.Com*. Retrieved 30 October 2007 from <http://www.Newscientist.com/article/dn12772-language-mutations-affect-leastused-words.html>.
- Holt, D. (1999a). *Pech (Paya)* (Vol. 366). München: Lincom Europa.
- Holt, D. (1999b). *Tol (Jicaque)* München: Lincom Europa.
- Holyoak, K. J., & Thagard, P. (1997). The analogical mind. *American psychologist*, 52(1), 35-44.
- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization* (2nd ed.). Cambridge: Cambridge University Press.
- Hori, H. (2001). Classifiers in Skidegate Haida. In O. Miyaoka & F. Endo (Eds.), *Languages of the North Pacific Rim* (Vol. 6). Osaka.
- Houle, D., & Kondrashov, A. (2003). Mutation [Electronic Version]. Retrieved 25 October 2007 from <http://bsc5936-03.fa04.fsu.edu/MUTATIONsubmit.pdf>
- Hualde, J. I. (1992). *Catalan: descriptive grammar*: Routledge.
- Hualde, J. I., & Ortiz de Urbina, J. (2003). *A grammar of Basque*. Berlin; New York: Mouton de Gruyter.

- Huffman, F. E. (1967). *An outline of Cambodian grammar*. Cornell University Ph.D. dissertation.
- Huffman, F. E. (1973). Thai and Cambodian: a case study of syntactic borrowing. *Journal of the American oriental society* 93(4), 488-509.
- Hurford, J. R. (1987). *Language and number: the emergence of a cognitive system*. Oxford: Basil Blackwell.
- Hurford, J. R. (1999). Artificially growing a numeral system. In J. Gvozdanović (Ed.), (pp.1-41).
- Hurford, J. R. (2001). Numeral systems. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (Vol.16, pp. 10756-10761). Oxford: Elsevier.
- Hurford, J. R. (2003). The interaction between numerals and nouns. In F. Plank (Ed.), *Noun phrase structure in the languages of Europe* (pp. 561-620). Berlin: Mouton de Gruyter.
- Hutt, M., & Subedi, A. (2003). *Nepali*. London: Teach Yourself.
- Huttar, G. L., & Huttar, M. L. (1994). *Ndyuka*. London: Routledge.
- Ikoro, S. M. (1994). Numeral classifiers in Kana. *Journal of African languages and linguistics*, 15, 7-28.
- Inkelas, S. (2006). Reduplication. In A. Anderson & E. K. Brown (Eds.), (pp. 417-419).
- Inkelas, S., & Zoll, C. (2005). *Reduplication: doubling in morphology*. Cambridge: Cambridge University Press.
- Innes, G. (1966). *An introduction to Grebo*: London: University of London, School of Oriental & African Studies.
- Iwasaki, S. (2002). *Japanese*. Amsterdam: John Benjamins.
- Jackson, K. H. (1953). *Language and history in early Britain: a chronological survey of the Brittonic languages, first to twelfth century A.D.* Edinburgh: Edinburgh University Press; Oliver & Boyd.
- Jacob, J. M. (1965). Notes on the numerals and numeral coefficients in Old, Middle and Modern Khmer. *Lingua*(15), 143-162.
- Jacob, J. M. (1990). *Introduction to Cambodian*. Oxford; New York: Oxford University Press.
- Jacq, P., & Sidwell, P. (1999). *Sapuan (Sepuar)*. München: Lincom Europa.
- Janhunen, J. (2003). *The Mongolic languages*. London: Routledge.
- Jespersen, O. (1922). *Language: its nature, development and origin*. London: G. Allen & Unwin ltd.
- Jones, R. B. (1970). Classifier constructions in Southeast Asia. *Journal of the American oriental society*, 90(1), 1-12.
- Jones, W., & Jones, P. (1991). *Barasano syntax*. Dallas: Summer Institute of Linguistics, University of Texas at Arlington.
- Joseph, B. D., & Janda, R. D. (Eds.). (2003). *The handbook of historical linguistics*. Malden, MA; Oxford: Blackwell.
- Joseph, B. D., & Philippaki-Warbuton, I. (1987). *Modern Greek*. London: Croom Helm.

- Kajitani, M. (2005). Semantic properties of reduplication among the world's languages [Electronic Version]. *LSO Working Papers in Linguistics 5: Proceedings of WIGL 2005* 93-106. Retrieved 5 December 2006 from <http://ling.wisc.edu/lso/wpl/5.1/LSOWP5.1-08-Kajitani.pdf>.
- Kakumasu, J. (1986). Urubu-Kaapor. In D. C. Derbyshire & G. K. Pullum (Eds.), *Handbook of Amazonian languages* (pp. 326-403). Berlin: Mouton de Gruyter.
- Kari, E. E. (1997). *Degema*. München ; Newcastle: LINCOM EUROPA.
- Kari, E. E. (2004). *A reference grammar of Degema*. Köln: Köppe.
- Kawasha, B. (2003). *Lunda grammar: a morphosyntactic and semantic analysis*. Ann Arbor: UMI.
- Kaye, A. S. (1987). Arabic. In B. Comrie (Ed.), *The world's major languages* (pp. 664-685). London: Croom Helm.
- Kenesei, I., Fenyvesi, A., & Vago, R. M. (1998). *Hungarian*. London: Routledge.
- Key, H. H. (1967). *Morphology of Cayuvava*. The Hague: Mouton.
- Kibrik, A. E., Kodzasov, S. V., & Muravyova, I. A. (2004). *Language and folklore of the Alutor people* Osaka ELPR
- Kimball, G. D. & Abbey, B. (1991). *Koasati grammar*. Lincoln: University of Nebraska Press.
- Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press.
- King, G. (1993). *Modern Welsh: a comprehensive grammar*. London: Routledge.
- Kinn, T. (2004). Demotion of numeral nouns from heads to dependents-a cognitive explanation [Electronic Version], abstract. Retrieved 5 June 2006 from <http://email.eva.mpg.de/~gil/numerals/abstracts/Kinn.pdf>.
- Klamer, M. A. F. (1998). *A grammar of Kambara*. Berlin: Mouton de Gruyter.
- Knowles, S. M. (2003). *A descriptive grammar of Chontal Maya*. Ann Arbor: UMI
- Koptjevskaja-Tamm, M. (2006). Partitives. In A. Anderson & E. K. Brown (Eds.), (pp. 218-221).
- Koptjevskaja-Tamm, M., & Wälchli, B. (2001). The Circum-Baltic Languages: An Areal-Typological Approach. In Ö.Dahl & M. Koptjevskaja-Tamm (Eds.), (pp. 615-750).
- Kornfilt, J. (1997). *Turkish*. London: Routledge.
- Kouwenberg, S. (1994). *A grammar of Berbice Dutch Creole* (Vol. 12). Berlin: Mouton de Gruyter
- Kozintseva, N. (1995). *Modern Eastern Armenian*. München: Lincom Europa.
- Kruspe, N. (2004). *A grammar of Semelai*. Cambridge: Cambridge University Press.
- Krute, L. D. (2003). *Piaroa nominal morphosemantics*. Ann Arbor: UMI
- Kuper, A. (2004). Primitive society. In A. Kuper & J. Kuper (Eds.), *The social science encyclopedia* (3rd ed., Vol. 2, pp. 800-802). London: Routledge.
- Kuryłowicz, J. (1965/1976). The evolution of grammatical categories. *Esquisses linguistiques*, 2(reprinted, 1976), 38-54.
- Kutsch Lojenga, C. (1994). *Ngiti: a Central-Sudanic language of Zaire*. Köln: R. Köppe.
- Lamb, W. (2001). *Scottish Gaelic*. München: Lincom Europa.
- Langacker, R. W. (1982). *Uto-Aztecan grammatical sketches*. Dallas: SIL.

- Langacker, R. W. (1987). *Foundations of cognitive grammar*. Stanford: Stanford University Press.
- LaPolla, R. J. (2003a). Dulong. In G. Thurgood & R. J. LaPolla (Eds.), (pp. 674-682).
- LaPolla, R. J. (2003b). Overview of Sino-Tibetan morphosyntax. In G. Thurgood & R. J. LaPolla (Eds.), (pp. 22-42).
- LaPolla, R. J. (2003c). Qiang. In G. L. Thurgood, R. J. (Ed.), (pp. 573-587).
- LaPolla, R. J. (with Huang, C.) (2003d). *A grammar of Qiang: with annotated texts and glossary*. Berlin: Mouton de Gruyter.
- Larsson, L.-G. (2001). Baltic influence on Finnic languages. In Ö. Dahl & M. Koptjevskaja-Tamm (Eds.), (pp. 237-256).
- Lass, R. (1990). How to do things with junk: exaptation in language change. *Journal of linguistics*, 26, 79-102.
- Le Page, R. B., & De Camp, D. (1960). *Jamaican Creole: an historical introduction to Jamaican Creole*. London: Macmillan: St. Martin's Press.
- Lehmann, C. (1995). *Thoughts on grammaticalization*. München: Lincom Europa.
- Lehmann, T. (1989). *A grammar of modern Tamil* (2nd ed.). Pondicherry: Pondicherry Institute of Linguistics and Culture.
- Lehmann, W. P. (1962). *Historical linguistics: an introduction*. London: Holt, Rinehart and Winston.
- Leopold, W. (1930). Polarity in language. *Language*, 6(4), 102-109.
- Levinton, J. S. (2001). *Genetics, paleontology, and macroevolution* (2nd ed.). Cambridge: Cambridge University Press.
- Li, C. N., & Thompson, S. A. (1977). The causative in Wappo: a special case of doubling. *Proceedings of the Third Annual Meeting of the Berkeley Linguistics Society, Berkeley*, 175-181.
- Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: a functional reference grammar*. Berkeley ; London: University of California Press.
- Lin, H. (2001). *A grammar of Mandarin Chinese*. München: Lincom Europa.
- Lindenfeld, J. (1973). *Yaqui syntax*. Berkeley: University of California Press.
- Lindström, E. (2002). *Topics in the grammar of Kuot: a non-Austronesian language of New Ireland, Papua New Guinea* from <http://www.ling.su.se/staff/evali/thesis/Kuot-PhD.html>.
- Linn, M. S. (2002). *A grammar of Euche (Yuchi)* Ann Arbor: UMI.
- Lorenzino, G. A. (1998). *The Angolar Creole Portugese of São Tomé: its grammar and sociolinguistic history* (Vol. 1). München: Lincom Europa
- Lorimer, D. L. R., & Morgenstierne, G. (1935). *The Burushaski language*. Oslo Cambridge, Mass.: Harvard University Press.
- Luján Martínez , E. R. (1999). The Indo-European system of numerals from '1' to '10'. In J. Gvozdanović (Ed.), (pp. 199-219).
- Lukas, J. (1937). *A study of the Kanuri language, grammar and vocabulary*. London, New York for the International Institute of African Languages & Cultures by the Oxford University Press.
- Lyman, T. A. (1979). *Grammar of Mong Njua (Green Miao): a descriptive study*. St. Helena, California: Lyman (published by the author).

- Lynch, J. (1998). *Pacific languages: an introduction*. Honolulu: University of Hawai'i Press.
- Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.
- Mac Eoin, G. (1993/2002). Irish. In M. J. Ball & J. Fife (Eds.), (pp. 101-144).
- MacKay, C. J. (1999). *A grammar of Misanla Totonac*. Salt Lake City: University of Utah Press.
- Mahootian, S., & Gebhardt, L. (1997). *Persian*. London: Routledge.
- Manoharan, S. (1989). *A descriptive and comparative study of Andamanese language*. Calcutta: Anthropological Survey of India.
- Manrique Castanade, L. (1967). Jilian Panpame. In N. A. McQuown & R. Wauchope (Eds.), *Handbook of Middle American Indians: Linguistics* (Vol. 5, pp. 331-348). Austin: University of Texas Press.
- Marcantonio, A. (2002). *The Uralic language family: facts, myths and statistics*. Oxford: Blackwell.
- Marsden, R. (2004). *The Cambridge Old English reader*. Cambridge: Cambridge University Press.
- Maslova, E. (2003a). *A grammar of Kolyma Yukaghir*. Berlin: Mouton de Gruyter.
- Maslova, E. (2003b). *Tundra Yukaghir*. München: Lincom Europa.
- Mather, E., Meade, M., & Miyaoka, O. (2002). *Survey of Yup'ik grammar*. Kyoto: ELPR.
- Matthews, P. H. (1997). *The concise Oxford dictionary of linguistics*. Oxford: Oxford University Press.
- Matthews, W. K. (1960). *Russian historical grammar*. London: University of London, Athlone Press.
- May, K. (1997). *A study of the Nimboran language: phonology, morphology and phrase structure*. Unpublished Thesis (M.A.) - La Trobe University.
- Mazaudon, M. (2003). Tamang. In G. Thurgood & R. J. LaPolla (Eds.), (pp. 291-314).
- McCone, K. (2005). *A first Old Irish grammar and reader (including an introduction to Middle Irish)*. Maynooth: National University of Ireland
- McDaniel, M. (2002). *English to Akha hand book 2000 words*. Retrieved 23 June 2008, from <http://www.akha.org/upload/akhabooks/english-to-akhahandbook.pdf>.
- McKay, G. (2000). Ndjébbana. In R. M. W. Dixon & B. J. Blake (Eds.), *The handbook of Australian languages* (pp. 154-354). Oxford: Oxford University Press.
- McMahon, A. M. S. (1994). *Understanding language change*. Cambridge: Cambridge University Press.
- Mel'chuk, I. A. (1988). *Dependency syntax: theory and practice*. Albany: State University Press of New York.
- Menninger, K. W., & Broner, P. (1969). *Number words and number symbols: a cultural history of numbers*. Cambridge: MIT Press.
- Miller, W. R. (1965). *Acoma grammar and texts*. Berkeley: University of California Press
- Miranda, R. V. (1975). Internal reconstruction: scope and limits. *Lingua*, 36(4), 289-305.
- Mithun, M. (1988). Lexical categories and the evolution of number marking. In M. Hammond & M. Noonan (Eds.), *Theoretical morphology: approaches in modern linguistics* (pp. 211-234). San Diego: Academic Press.

- Mithun, M. (1999). *The languages of native North America*. Cambridge: Cambridge University Press.
- Morpurgo Davies, A. (1987). 'Organic' and 'organism' in Franz Bopp. In H. M. Hoenigswald & L. F. Wiener (Eds.), *Biological metaphor and cladistic classification an interdisciplinary perspective* (pp. 91-107). London: Pinter.
- Mossâe, F. (1968). *A handbook of Middle English*. Baltimore: John Hopkins Press.
- Muntzel, M. C. (1986/2003). *The structure of Ocuilteco*. Ann Arbor: UMI.
- Mustanoja, T. F. (1960). *A Middle English syntax*: Helsinki.
- Nagaraja, K. S. (1999). *Korku language: grammar, texts, and vocabulary*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Naish, C. M. (1979). A syntactic study of Tlingit [microform]. On *Language data. Amerindian series; no. 6*. Dallas: SIL.
- Nau, N. (1998). *Latvian*. München: Lincom Europa.
- Neidle, C. J. (1988). *The role of case in Russian syntax*. Kluwer Academic Publishers, Dordrecht; Boston.
- Nelson, D. (1998). *Grammatical case assignment in Finnish*. New York: Garland Pub.
- Nelson, D., & Toivonen, I. (2000). Counting and the grammar: case and numerals in Inari Sami [Electronic Version]. *Leeds Working Papers in Linguistics* 179-192. Retrieved 25 October 2007 from <http://www.leeds.ac.uk/linguistics/WPL/WP2000/Nels&Toiv2.pdf>.
- Neukom, L. (2001). *Santali*. Unpublished Thesis (Ph.D.)-Universität, Zürich.
- Nevins, A., Pesetsky, D., & Rodrigues, C. (2007). Pirahã exceptionality: a reassessment [Electronic Version], 1-58. Retrieved 26 December 2007 from http://www.linguistics.uiuc.edu/LSO/Critique_of_Everett.pdf
- Newmark, L. (1957). *Structural grammar of Albanian*. Indiana University: Bloomington
- Newmark, L., Hubbard, P., & Prifti, P. R. (1982). *Standard Albanian: a reference grammar for students*. Stanford, California: Stanford University Press.
- Newmeyer, F., J. (2002). Uniformitarian assumption and language evolution research. In A. Wray (Ed.), *The transition to language* (pp. 359-375). Oxford: Oxford University Press.
- Ngom, F. (2003). *Wolof*. München: Lincom Europa.
- Nguyen, T. H. (2004). *The structure of the Vietnamese noun phrase*. Unpublished Thesis (Ph.D.)-Boston University.
- Nichols, L. (1997/2003). *Topics in Zuni syntax*. Ann Arbor: UMI.
- Nikolaeva, I. A., & Tolskaya, M. (2001). *A grammar of Udihe*. Berlin: Mouton de Gruyter.
- Noonan, M. (1992). *A grammar of Lango*. Berlin: Mouton de Gruyter.
- Noonan, M. (2003). Chantyal. In G. Thurgood & R. J. LaPolla (Eds.), (pp. 315-335).
- Noonan, M., Bhulanja, R. P., Chantyal, J. M., & Pagliuca, W. (1999). *Chantyal dictionary and texts*. Berlin ; New York: Mouton de Gruyter.
- Nordlinger, R. (1998). *A grammar of Wambaya, Northern Australia*. Canberra: Pacific Linguistics.
- Norman, J. (1988). *Chinese*. Cambridge: Cambridge University Press.

- Ò Dochartaigh, C. (1992). The Irish Language. In D. MacAulay (Ed.), *The Celtic languages* (pp. 11-99). Cambridge: Cambridge University Press.
- Ò Siadhail, M. (1989). *Modern Irish: grammatical structure and dialectal variation*. Cambridge: Cambridge University Press.
- Odé, C. (2002). A Sketch of Mpur. In G. P. Reesink (Ed.), *Languages of the eastern Bird's Head Pacific linguistics; 524*. (pp. 45-107). Canberra: Australian National University.
- Ojeda, A. E. (1994). Number invariance: summary [Electronic Version], 1. Retrieved 5 October 2007 from <http://www.linguistlist.org/issues/5/5-1362.html>
- Olawsky, K. J. (2002). *Urarina texts*. München: Lincom Europa.
- Orwin, M. (1995). *Colloquial Somali: a complete language course*. London: Routledge.
- Osborne, C. R. (1974). *The Tiwi language: grammar, myths and dictionary of the Tiwi language spoken on Melville and Bathurst Islands, northern Australia*. Canberra: Australian Institute of Aboriginal Studies.
- Patrick, P. L. (2003). Jamaican Creole morphology and syntax [Electronic Version]. Retrieved 3 July 2008 from <http://privatewww.essex.ac.uk/~patrickp/papers/JamCreoleGrammar.pdf>.
- Payne, D. L. (2003). *Aspects of the grammar of Yagua: a typological approach*. Ann Arbor: UMI.
- Peeke, C. (1994). Waorani. In P. Kahrel & R. v. d. Berg (Eds.), *Typological studies in negation* (pp. 267-290). Amsterdam: John Benjamins.
- Penchoen, T. G. (1973). *Tamazight of the Ayt Ndhir*. Los Angeles: Undena Publications.
- Peyraube, A. (2004). Ancient Chinese. In R. D. Woodard (Ed.), *The Cambridge encyclopedia of the world's ancient languages* (pp. 988-1014). Cambridge: Cambridge University Press.
- Phythian, J. E. Counting Systems of Papua New Guinea: An appreciation of the work of the later Dr. Glendon Lean [Electronic Version], 1-15. Retrieved 31 July 2007.
- Plaisier, H. (2003). Lepcha. In G. Thurgood & R. J. LaPolla (Eds.), (pp. 705-716).
- Plungian, V. (1995). *Dogon*. München; Newcastle: Lincom Europa.
- Premisrirat, S. (1987). *Khmu, a minority language of Thailand*. Canberra: Australian National University.
- Press, I. (1986). *A grammar of modern Breton*. Berlin: Mouton de Gruyter.
- Procházka, S. (2006). Arabic. In E. K. Brown and A. Anderson (Eds.), (pp. 423-431).
- Quesada, J. D. (2000). *A grammar of Teribe*. München: Lincom Europa.
- Quirk, R., & Wrenn, C. L. (1957). *An Old English grammar* (2nd ed.). London: Methuen.
- Rajam, V. S. (1992). *A reference grammar of classical Tamil poetry: 150 B.C. - pre-fifth/sixth century A.D.* Philadelphia: American Philosophical Society.
- Ramaswami, N. (1992). *Bhumij grammar*. Mysore: Central Institute of Indian Languages.
- Rau, D.-H. V. (1992). *A grammar of Atayal*. Ann Arbor: UMI
- Reesink, G. P. (1999). *A grammar of Hatam, Bird's Head Peninsula, Irian Jaya*. Canberra: Australian National University.

- Reesink, G. P. (2002). *Languages of the eastern Bird's Head*. Canberra: Pacific Linguistics Research School of Pacific and Asian Studies Australian National University.
- Rennison, J. R. (1997). *Koromfe*. London: Routledge.
- Riccardi, T. (2003). Nepali. In G. Cardona & D. Jain (Eds.), (pp. 538-580).
- Rijkhoff, J. (2002). *The noun phrase*. Oxford: Oxford University Press.
- Ringe, D., Warnow, T., & Taylor, A. (2002). Indo-European and Computational Cladistics [Electronic Version]. *Transactions of the Philological Society*, 100, 59-29 from <http://www.cs.rice.edu/~nakhleh/CPHL/RWT02.pdf>.
- Roberts, J. R. (1987). *Amele*. London: Croom Helm.
- Robins, R. H. (1958). *The Yurok language: grammar, texts, lexicon*. Berkeley: University of California Press.
- Rodrigues, A. D. (1999). Macro-Jê. In R. M. W. Dixon & A. Aikhenvald (Eds.), (pp. 165-206). Cambridge: Cambridge University Press.
- Romaine, S. (2006). Pidgins and creoles: overview. In A. Anderson & E. K. Brown (Eds.), (pp. 600-606).
- Romero-Figeroa, A. (1997). *A reference grammar of Warao*. München: Lincom Europa.
- Rood, D. S., & Charney, J. O. (1976). *Wichita grammar*. New York: Garland.
- Ross, M. (1980). Some elements of Vanimo, a New Guinea tone language. In D. J. P. K. G. Holzkecht (Ed.), *Papers in New Guinea Linguistics* (Vol. 20, pp. 77 - 109). Canberra.
- Rowlands, E. C. (1993). *Yoruba*. Sevenoaks: Teach Yourself Books.
- Rubino, C. (2005). Reduplication. In M. Haspelmath, M. Dryer, D. Gil & B. Comrie (Eds.), (pp. 114-117).
- Rude, N. E. (1985). *Studies in Nez Perce grammar and discourse*. Ann Arbor: UMI
- Ruhlen, M. (1987). *A guide to the world's languages*. Stanford: Stanford University Press.
- Rumsey, A. (2000). Bunuba. In R. M. W. Dixon & B. J. Blake (Eds.). *The handbook of Australian languages* (pp. 34-152). Oxford: Oxford University Press.
- Saeed, J. I. (1993). *Somali reference grammar*. Kensington, Md.: Dunwoody Press.
- Saeed, J. I. (1999). *Somali*. Amsterdam: John Benjamins.
- Sakel, J. (2003). *A grammar of Mosenen*. Unpublished Thesis (Ph.D)-Katholieke Universiteit Nijmegen.
- Sanches, M., & Slobin, L. (1973). Numeral classifiers and plural marking: an implicational universal. *Working papers in language universals*, 1-22.
- Schadeberg, T. C., Elias, P., & Muratori, C. (1979). *A description of the Orig language (Southern Kordofan)*. Tervuren: Museum royal de l'Afrique centrale.
- Schmandt-Besserat, D. (1996). *How writing came about*. Austin: University of Texas Press.
- Schuh, R. G. (1999). Hausa grammar online [Electronic Version]. Retrieved 14 January 2006 from <http://www.humnet.ucla.edu/humnet/aflang/Hausa/indexframe.html>.
- Seiler, W. (1985). *Imonda, a Papuan language*. Canberra.
- Senft, G. (1996). *Classificatory particles in Kilivila*. Oxford: Oxford University Press.
- Shafer, R. (1941). Nahali: A Linguistic Study on Paleoethnography. *Harvard Journal of Asiatic Studies*, 5(3/4.), 346-371.

- Shlonsky, U. (2004). The form of Semitic noun phrases. *Lingua*, 114 (12), 1465-1526.
- Shnukal, A. (1988). *Broken: an introduction to the Creole language of Torres Strait*. Canberra: Australian National University
- Sidwell, P. (1999). The Austro-Asiatic numerals '1' to '10' from a historical and typological perspective. In J. Gvozdanović (Ed.), (pp. 253-271).
- Siewierska, A. (2005). Gender distinctions in independent personal pronouns. In M. Haspelmath, M. Dryer, D. Gil & B. Comrie (Eds.), (pp. 182-185).
- Sigler, M. (1992). Number agreement and specificity in Armenian. In C. P. Canakis, G. P. Chan & J. M. Denton (Eds.), *CLS 28: papers from the 28th regional meeting of the Chicago Linguistic Society, 1992* (pp. 499-514). Chicago: Chicago Linguistic Society.
- Simon, T., Hespos, S., & Rochat, P. (1995). Do infants understand simple arithmetic? A replication of Wynn (1992). *Cognitive Development*, 10, 253-269.
- Simpson, A. (2005). Analyticity in the nominal domain: classifiers and the structure of DPs in languages of Southeast Asia [Electronic Version], 1-24. Retrieved 25 May 2007 from http://www.people.fas.harvard.edu/~ctjhuang/Course_LSA222/handouts/Simpson1a.pdf.
- Sinha, N. K. (1975). *Mundari grammar*. Mysore: Central Institute of Indian Languages.
- Slater, K. W. (2003). *A grammar of Mangghuer: a Mongolic language of China's Qinghai-Gansu Sprachbund*. London: Routledge Curzon.
- Smirnova, M. (1982). *The Hausa language: a descriptive grammar*. London: Routledge & Kegan Paul.
- Smith, J. M., & Szathmáry, E. (1999). *The origins of life: from the birth of life to the origin of language*. Oxford: Oxford University Press.
- Sohn, H.-M. (1994). *Korean*. London: Routledge.
- Sohn, H.-M. (1999). *The Korean language*. Cambridge: Cambridge University Press.
- Staley, W. E. (2003). *Referent management in Olo: a cognitive perspective*. Ann Arbor: UMI.
- Steever, S. B. (1998). *The Dravidian languages*. London: Routledge.
- Stevick, E. W. (2003). *Syntax of colloquial East Armenian*. Ann Arbor: UMI
- Stevick, R. D. (1963). The biological model and historical linguistics. *Language*, 39(2), 159-169.
- Stolz, T., & Veselinova, L. (2005). Ordinal numerals. In M. Haspelmath, M. Dryer, D. Gil & B. Comrie (Eds.), (pp. 218-221).
- Storch, A. (1997). Where have all the noun classes gone? A case study of Jukun. *Journal of African languages and linguistics*, 18, 157-170.
- Sulkala, H., & Karjalainen, M. (1992). *Finnish*. London: Routledge.
- Suttles, W. P. (2004). *Musqueam reference grammar*. Vancouver: University of British Columbia Press.
- Tamura, S. (2000). *The Ainu language* (Vol. 2). Tokyo: Sanseido
- Tao, L. (2005). The importance of discourse analysis for linguistic theory: a Mandarin Chinese illustration. In Z. Frajzyngier, A. Hodges & D. S. Rood (Eds.), *Linguistic diversity and language theories* (Vol. 72, pp. 285-317). Amsterdam: John Benjamins.

- Teferra, A. (1991). Sketch of Shabo Grammar. In M. L. Bender (Ed.), *Proceedings of the Fourth Nilo-Saharan Conference, Bayreuth, Aug. 30-Sep. 2, 1989* (pp. 371-381). Hamburg: H. Buske.
- Tenser, A. (2005). *Lithuanian Romani*. München: Lincom Europa.
- Terrill, A. (2003). *A grammar of Lavukaleve*. Berlin: Mouton de Gruyter.
- Tharp, D. (1996). Sulka grammar essentials. In J. M. Clifton (Ed.), *Two Non-Austronesian grammars from the islands* (pp. 77-179). Ukarumpa, EHP, Papua New Guinea: SIL.
- Thomas, D. D. (1971). *Chrau grammar*. Honolulu: University of Hawaii Press.
- Thomason, S. G. (2001). *Language contact*. Edinburgh: Edinburgh University Press.
- Thomason, S. G. (2006). Syntactic reconstruction. In A. Anderson & E. K. Brown (Eds.), (pp. 397-400).
- Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Thorne, D. A. (1993). *A comprehensive Welsh grammar*. Oxford: Blackwell.
- Thurgood, G., & LaPolla, R. J. (2003). *The Sino-Tibetan languages*. New York: Routledge.
- Thurneysen, R. (1946). *A grammar of Old Irish* Dublin: The Dublin Institute for Advanced Studies
- Triulzi, A., A.A. Dafallah, and M.L. Bender. (1976). In M. L. Bender (Ed.), *The Non-Semitic languages of Ethiopia* (pp. 513-532). East Lansing: African Studies Center, Michigan State University.
- Tucker, A. N., & Bryan, M. A. (1966). *Linguistic analyses: the non-Bantu languages of North-Eastern Africa*. London: Oxford University Press.
- Turnbull, A., & Kilgour, R. (1923). *Nepali grammar and vocabulary* (3rd ed.). London: W. Thacker.
- Vajda, E. J. (2004). *Ket*. München: Lincom Europa.
- Vamarasi, M. K. (2002). *Rotuman*. München: Lincom Europa.
- van Staden, M. (2000). *Tidore: a linguistic description of a language of the North Moluccas*. Leiden: Universiteit Leiden.
- Velupillai, V. (2003). *Hawai'i Creole English: a typological analysis of the tense-mood-aspect system* Houndmills [u.a.] Palgrave Macmillan
- Verhaar, J. W. M. (1995). *Toward a reference grammar of Tok Pisin: an experiment in corpus linguistics*. Honolulu: University of Hawaii Press.
- Vidal, A. (2002). *Pilagá grammar (Guaykuruan family, Argentina)* Ann Arbor: UMI
- Vlasto, A. P. (1986). *A linguistic history of Russia to the end of the eighteenth century*. Oxford: Clarendon.
- Volker, C. A. (1998). *The Nalik language of New Ireland, Papua New Guinea*. New York: Peter Lang.
- Voort, H. (2004). *A grammar of Kwaza*. Berlin ; New York: Mouton de Gruyter.
- Wang, L. (1994). *Origin and development of classifiers in Chinese*. The Ohio State University.
- Wang, L. (1996). Origin of Classifiers in Proto-Chinese [Electronic Version], 65-106. Retrieved 17 April 2007 from <http://pears.lib.ohio-state.edu/China/origin.doc>.

- Watkins, T. A. (1993). Welsh. In J. Fife & M. J. Ball (Eds.), (pp. 289-348). London ; New York : Routledge.
- Watters, J. R. (1981). *A phonology and morphology of Ejagham (with notes on dialect variation)*. Unpublished Thesis (Ph.D.)- University of California, Los Angeles.
- Weber, D. (1989). *A grammar of Huallaga (Huánuco) Quechua*. Berkeley: University of California Press.
- Wellens, I. H. W. (2003). *An Arabic creole in Africa: the Nubi language of Uganda* Unpublished Thesis (Ph.D)- Nijmegen University.
- Welmers, W. E. (1973). *African language structures*. Berkeley: University of California Press.
- Wiersma, G. (2003). Yunnan Bai. In G. Thurgood & R. J. LaPolla (Eds.), (pp. 651-673).
- Wiese, H. (2007). The co-evolution of number concepts and counting words. *Lingua*, 117(5), 758-772.
- Williamson, J. S. (2003). *Studies in Lakhota grammar*. Ann Arbor: UMI.
- Williamson, K. (1965). *A Grammar of the Kolokuma Dialect of Ijo*. University Press: Cambridge.
- Williamson, K., & Blench, R. (2000). Niger-Congo. In B. Heine & D. Nurse (Eds.), *African languages: an introduction* (pp. 11-42).
- Wilson, J. P. (1996). *Binandere nominal structures*. Unpublished Thesis (M.A.)- University of Texas at Arlington.
- Wilson, J. P. (2002). *Binandere verbal structures*. Retrieved 29 June 2008, from http://www.sil.org/pacific/png/pubs/0000015/Binandere_Grammar_Essentials.pdf
- Wilson, P. R. (1980). *Ambulas Grammar*. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.
- Winter, W. (1999). When numeral systems are expanded. In J. Gvozdanović (Ed.), (pp. 43-53).
- Wischer, I. (2006). Grammaticalization. In A. Anderson & E. K. Brown (Eds.), (pp. 129-136).
- Wolfart, H. C., & Carroll, J. F. (1981). *Meet Cree: a guide to the Cree language*. Lincoln: University of Nebraska Press.
- Wurm, S. A. (1997). *Materials on languages in danger of disappearing in the Asia-Pacific region No. 1: some endangered languages of Papua New Guinea: Kaki Ae, Musom, and Aribwatsa*. Canberra: Pacific Linguistics Research School of Pacific and Asian Studies Australian National University.
- Yashwanta Singh, C. (2000). *Manipuri grammar*. New Delhi: Rajesh Publications.
- Yasugi, Y. (1995). *Native Middle American languages: an areal-typological perspective*. Osaka: National Museum of Ethnology.
- Yumitani, Y. (2004). *A phonology and morphology of Jemez Towa* Ann Arbor: UMI.
- Zabbal, Y. (2002). *The semantics of number in the Arabic noun phrase*. Unpublished Thesis (M.A.)-University of Calgary, Alberta.
- Zavala, R. (2002). *Inversion and other topics in the grammar of Olutec (Mixean)* Ann Arbor: UMI
- Zipf, G. K. (1949). *Human behaviour and the principle of least effort: an introduction to human ecology*. Cambridge, Mass.: Addison-Wesley Press.
- Zúñiga, F. (2000). *Mapudungun*. München: Lincom Europa.

Zylstra, C. F. (1991). A syntactic sketch of Alacatzala Mixtec. In C. H. Bradley & B. E. Hollenbach (Eds.), *Studies in the syntax of Mixtecan languages* (Vol. 3, pp. 1-178). Dallas: SIL.