LOSS, RETENTION, AND ACCRETION OF INFLEXIONAL CASE-MARKING - DIACHRONIC EVIDENCE FROM ITALIC AND FINNIC.

Zeprina-Jaz Ainsworth

Worcester College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Summer 2019.

ABSTRACT OF THE DISSERTATION

‘Loss, Retention, and Accretion of Inflexional Case-Marking - Diachronic Evidence from Italic and Finnic’

Zeprina-Jaz Ainsworth, Worcester College

Doctoral thesis submitted for DPhil General Linguistics and Comparative Philology, Summer 2019.

The development of inflexional case systems has long been of interest to historical linguists. Languages such as Latin, Ancient Greek, and many other Indo-European varieties have lost morphological case distinctions through time. Cross-linguistically, however, we find quite different changes affecting case systems. For instance, Finnish has largely maintained the case system reconstructed for proto-Finnic, whereas the related variety Veps has innovated novel inflexional case-markers.

Although much work has been done to explain the changes that are found in individual varieties, a cross-linguistic comparison of the types of development that are found - that is, the loss, maintenance, or propagation of inflexional case systems - which aims to understand *why* these particular changes occur has yet to be attempted. This study offers a comparison of several varieties from two unrelated languages families: Italic (Latin and the Romance languages) and Finnic. An understanding of the different morphological structures found in these varieties and how these structures are exploited by language-users in the production of novel inflected forms offers insight not only into *what* can happen in the inflexional case system diachronically, but *why* these systems have developed in different ways. Reference to a broader range of languages confirms

that there is an implicational relationship between the types of morphological structure found in a language and the size of the inflexional case system.

CONTENTS

Acknowledgements 9

List of figures 11

List of tables 12

List of abbreviations 17

1. INTRODUCTORY CONSIDERATIONS

1.0 THE TOPIC 18

1.1.0 AN EVOLUTIONARY PERSPECTIVE 21

1.1.1 *Diachrony and the Language-User* 24

1.1.2 *Patterns in Morphology* 25

1.1.3 *A Language-User Perspective on 'Case'* 31

1.2.0 THE DATA SET 35

1.2.1 *The Italic Languages* 36

1.2.2 *The Finnic Languages* 39

1.3 CLOSING REMARKS 43

2. THE ITALIC LANGUAGES

2.0 THE ITALIC CASE SYSTEM 45

2.1.0 INTRODUCTION TO LATIN 45

2.1.1 *The Latin Case System* 45

2.1.2 *The Development of the Latin Case System - Phonology* 49

2.1.3 *The Development of the Latin Case System - Analogy* 52

2.1.4 *The Development of the Italic Case Systems - Syntax* 56

2.1.5 *The Development of the Italic Case Systems - Contact Considerations* 61

2.2.0 INTRODUCTION TO OLD GALLO-ROMANCE 67

2.2.1 *The Old Gallo-Romance Case 'System'* 67

2.2.2 *The Development of the Old Gallo-Romance Case-Forms - Phonology* 73

2.2.3 *The Development of the Old Gallo-Romance Case-Forms - Analogy* 76

2.3.0 INTRODUCTION TO ROMANIAN 81

2.3.1 *The Romanian Case System* 81

2.3.2 *The Development of the Romanian Case System - Phonology* 85

2.3.3. *The Development of the Romanian Case System - Analogy* 88

2.4 CLOSING REMARKS 90

3. THE FINNIC LANGUAGES

3.0 THE FINNIC CASE SYSTEM 92

3.1.0 INTRODUCTION TO FINNISH 93

3.1.1 *The Finnish Case System* 93

3.1.2 *The Development of the Finnish Case System - Phonology* 102

3.1.3 *The Development of the Finnish Case System - Analogy* 106

3.1.4 *The Development of the Finnish Case System - Syntax* 110

3.1.5 *The Development of the Finnish Case System - Contact Considerations* 113

3.2.0 INTRODUCTION TO THE LIVONIAN LANGUAGE 117

3.2.1 *The Livonian Case System* 118

3.2.2 *The Development of the Livonian Case System - Phonology* 124

3.2.3 *The Development of the Livonian Case System - Analogy* 127

3.2.4 *The Development of the Livonian Case System - Syntax* 129

3.2.5 *The Development of the Livonian Case System - Contact Considerations* 131

3.3.0 INTRODUCTION TO THE VEPS LANGUAGE 133

3.3.1 *The Veps Case System* 134

3.3.2 *The Development of the Veps Case System - Phonology* 142

3.3.3 *The Development of the Veps Case System - Analogy* 146

3.3.4 *The Development of the Veps Case System - Syntax* 148

3.3.5 *The Development of the Veps Case System - Contact Considerations* 151

3.4 CLOSING REMARKS 154

4. THEORETICAL APPROACHES TO THE ITALIC AND FINNIC DATA

4.0 INTRODUCTION 156

4.1.0 MODELS OF MORPHOLOGICAL DESCRIPTION 156

4.1.1 '*Constructivist' Approaches* 158

4.1.2 *'Abstractivist' Approaches* 160

4.1.3 *Closing Remarks* 170

4.2.0 THE ITALIC LANGUAGES 170

4.2.1 *'(De-)Compositionality' and Attendant Issues* 171

4.2.2 *Allomorphy* 178

4.2.3 *Syncretism* 185

4.2.4 *Closing Remarks* 189

4.3.0 THE FINNIC DATA 189

4.3.1 *'(De-)Compositionality' and Attendant issues* 190

4.3.2 *Allomorphy* 194

4.3.3 *Syncretism* 205

4.3.4 *Closing Remarks* 209

4.4. CONSTRUCTIVIST AND ABSTRACTIVIST APPROACHES - CLOSING REMARKS 209

5. 'SUBSTANCE ENTITIES' AND 'SCHEMATIC PATTERNS'

5.0 INTRODUCTION 211

5.1.0 'SUBSTANCE ENTITIES' AND 'SCHEMATIC PATTERNS' 213

5.1.1 *Substance Entities, Schematic Patterns, and Reference to the* Paradigm 220

5.2 SYMBOLIC AND INDEXICAL INFORMATION 225

5.3.0 EVIDENCE FOR SUBSTANCE ENTITIES 229

5.3.1 *Evidence from L1 Learners of Finnish* 229

5.3.2 *Evidence from L2 Learners and Native Speakers of Finnish* 234

5.4 EVIDENCE FOR SCHEMATIC PATTERNS 239

5.5 EVIDENCE FOR PARADIGMATIC SCHEMATA 241

5.6 CLOSING REMARKS 243

6. THE DEVELOPMENT OF THE ITALIC AND FINNIC CASE SYSTEMS

6.0 SUBSTANCE ENTITIES, SCHEMATIC PATTERNS, AND THE DEVELOPMENT OF CASE SYSTEMS 244

6.1 MORPHOLOGICAL PATTERNS 245

6.2 CASE-REDUCING LANGUAGES 253

6.3 CASE-STEADY LANGUAGES 268

6.4 CASE-INCREASING LANGUAGES 274

6.5 CLOSING REMARKS 278

7. CROSS-LINGUISTIC SURVEY

7.0 INTRODUCTION 282

7.1.0 THE CROSS-LINGUISTIC DATA 283

7.1.1 *Languages with 6-7 Morphological Case-Forms* 288

7.1.2 *Languages with 8-9 Morphological Case-Forms* 289

7.1.3 *Languages with 10+ Morphological Case-Forms* 294

7.1.4 *Closing Remarks* 300

7.2.0 EXCEPTIONS 302

7.2.1 *The Selkup Case System* 302

7.2.2 *The Romance Verb and its Development* 308

7.2.3 *Analysis* 316

7.3 CLOSING REMARKS 327

8. CONCLUSIONS

8.0 CONCLUSIONS 329

References 331

ACKNOWLEDGEMENTS

This thesis owes a great deal to a great many people, and represents the supportive efforts of a host of friends, family members, peers, and academics. In the first place, I would like to thank my supervisors, Professor Martin Maiden and Professor Ulla-Maija Forsberg, for their invaluable advice and guidance throughout my studies, supporting the writing of this thesis as well as other projects undertaken alongside it.

Second, to my parents, whose financial aid was the least of the help offered in this endeavour: thank you for being there to bounce ideas off of, for challenging the most basic of my assumptions, and for the innumerable opportunities you have provided me throughout my academic career. I would never have been able to achieve half of what I have done without your unceasing love and support.

Third, I am grateful for the help offered by a wider network of colleagues and peers, who facilitated the growth of many of the ideas presented herein. I would particularly like to thank Professor Aditi Lahiri and Professor Wolfgang de Melo, who have offered valuable academic feedback and more general support throughout my studies. My warmest thanks also to Dr. Sandra Paoli and Dr. Chiara Cappellaro for their insightful comments on earlier drafts of work presented at my Transfer of Status and Confirmation of Status.

Fourth, I am extremely grateful to the administrative staff in the Faculty of Linguistics, Philology and Phonetics, whose hard work behind the scenes deserves a great deal more attention than it sometimes gets! Thank you particularly to Jane Cunning, Sarah Wild, and Dan Holloway.

Finally, this thesis is dedicated to my wonderful, supportive husband Rob, who has survived this DPhil with a good deal more grace and patience than myself. Thank you for

being there throughout to share in the excitement, the hard work, and the stress of this venture. This would not have been possible without you.

The financial support of the Birla and Kumar Studentship and the Vice-Chancellor's Fund is gratefully acknowledged. Any errors that remain in this thesis are entirely my own.

LIST OF FIGURES

Figure 2.1: Analogical development of non-etymological old Occitan nominative singular *paires* 78

Figure 3.1: Contact-induced analogical change of Livonian *-ks* 133

Figure 4.1: Purely formal four-part analogy of Latin {TEMPLUM} ‘temple’ 161

Figure 4.2: Calculating entropy 162

Figure 4.3: Calculating the entropy of a coin toss 163

Figure 4.4: Calculating conditional entropy 163

Figure 4.5: The analogical development of Romanian genitive-dative singular *grele* ‘hard’ 177

Figure 4.6: Decomposing Finnish comitative *kuivine* ‘dry’ 205

Figure 6.1: Visual representations of morphological patterns 245

Figure 6.2: Implicational relations in the Latin case system 249

Figure 6.3: Potential inflexional errors in Latin {MANUS} ‘hand’ 249

Figure 6.4: Implicational relations in Finnish *s*-stem nominals 250

Figure 6.5: Potential inflexional errors in the lexeme {UROS} ‘male; man’ 250

Figure 6.6: Multiple loci of informativeness in the Finnish case system 251

Figure 6.7: Analysing Latin nominative singular RĒX 254

Figure 6.8: The merging of relations between Veps word-level substance entities 275

Figure 7.1: The analogical deduction of Latin past passive participle forms 324

LIST OF TABLES

Table 1.1: General functions of the Finnic local cases 41

Table 1.2: Inflexion classes in Latin 42

Table 1.3: Inflexion classes in Finnish 42

Table 2.1: Major declensions in Latin 46

Table 2.2: First and second declensions Ancient Greek loanwords 47

Table 2.3: Major adjective declensions in Latin 47

Table 2.4: Loss of *-m* in Latin first declension singular nominals 50

Table 2.5: Phonological developments affecting first declension singular nominals 51

Table 2.6: Partial Latin second and fourth declension plural inflexional patterns 54

Table 2.7: Latin genitive and French *de* compared 59

Table 2.8: Functions of the old Gallo-Romance nominative and oblique cases 68

Table 2.9: Old French and old Occitan masculine nominal paradigms 70

Table 2.10: Old French {ARC}, {CHEVAUS}, and {GAS} fully declined 71

Table 2.11: Old French and old Occitan feminine nominal paradigms 71

Table 2.12: Old French and old Occitan gender-distinguishing adjectives 72

Table 2.13: Old Gallo-Romance {MURS} from Latin {MŪRUS} ‘wall’ 73

Table 2.14: Old French {LERRE} ‘thief’ from Latin {LATRŌ} ‘mercenary soldier’ 75

Table 2.15: Selected Romanian lexemes fully declined (non-definite) 82

Table 2.16: Inflexional patterns exhibited by Romanian feminine nouns 83

Table 2.17: Patterns of syncretism in Romanian feminine nominals 83

Table 2.18: Romanian definite declensions 85

Table 2.19: Romanian {LEGE} ‘law’ from Latin third declension (feminine) {LĒX} 86

Table 3.1: Finnish {PUHELIN} and {AJATUS} fully declined 94

Table 3.2: The functions of the ‘semantic’ Finnish cases 95

Table 3.3: Finnish declensions 98

Table 3.4: Distribution of case exponents in Finnish 100

Table 3.5: The traditional view of the development of the Finnic local case-forms 107

Table 3.6: The analogical development of the excessive case-form 107

Table 3.7: *s*-stem nominals in Finnish 108

Table 3.8: Livonian {LÄPŠ} and {KALĀ} fully declined 118

Table 3.9: Comparing weak and strong stems in selected Livonian nominals 121

Table 3.10: Livonian case exponents 121

Table 3.11: Livonian nominals {KALĀ}, {TUBĀ}, and {AIGĀ} fully declined 123

Table 3.12: Paradigm template for *kalā-*, *tubā-*, and *aigā*-type nominals 123

Table 3.13: Alternating forms in selected Livonian nominals 127

Table 3.14: {AMĀ} and {IZĀ} fully declined in the plural 128

Table 3.15: Inflexional patterns in {ĀIGA}, {SĪLMA}, {KĀNGA}, and {ĀRGA} plural forms 128

Table 3.16: Veps {SUGA} and {KÄZI} fully declined 135

Table 3.17: Veps inflexion classes 138

Table 3.18: Case suffixes in the three major Veps dialects 140

Table 3.19: Partitive exponents in Veps 141

Table 3.20: Distribution of illative singular exponents in Veps 141

Table 3.21: Approximative local case-forms across the Veps dialects 144

Table 3.22: The development of new inflexional case-forms in Veps 149

Table 3.23: Comparing morphological case-forms and postpositions in Veps 150

Table 3.24: The similarities between the Livonian and Veps contact situations 153

Table 4.1: Theories of inflexional morphology 157

Table 4.2: Modern Greek nominal inflexion classes 164

Table 4.3: Conditional entropies for Modern Greek paradigms in table 4.2 164

Table 4.4: Decomposing the singular forms of
Latin first declension {PUELLA} 172

Table 4.5: Decomposing the singular forms of {PUELLA} 172

Table 4.6: Analysing the plural forms of {PUELLA} 173

Table 4.7: Comparing first declension case morphs across singular and plural 175

Table 4.8: Morphological alternations in Romanian {ZI} and {STEA} 176

Table 4.9: Imparisyllabic forms in old French and old Occitan nouns 179

Table 4.10: Stem allomorphy in Latin {RĒX}, {URBS}, and {NOX} 181

Table 4.11: Stem allomorphy in Latin {SUPELLEX} 181

Table 4.12: Formal variation of the suffix in the first, second, and
third declensions in Latin 183

Table 4.13: Syncretism in (most) Romanian feminine nouns 185

Table 4.14: Selected old Occitan nominal paradigms 187

Table 4.15: Chiastic distribution of forms in *murs*-type nominals
(old French and old Occitan) 187

Table 4.16: Decomposing inflected forms of Veps {HEBO} 'horse' 190

Table 4.17: Cumulative morphs in Livonian plural forms 191

Table 4.18: Illative singular forms in written Veps 195

Table 4.19: Nominative and illative singular forms of selected Veps nominals 196

Table 4.20: Partial paradigms of selected Veps nominals 196

Table 4.21: Incorrect plural forms in of Finnish {KUIVA} 'dry' 199

Table 4.22: Morphophonological alternations in Livonian nominal stems 201

Table 4.23: Stem differences in *ajatus* and *taivas*-type nominals 203

Table 4.24: {UROS} 'man' partially declined 203

Table 4.25: Stem allomorphy distribution in Veps nominals 204

Table 4.26: Nominative singular and genitive singular forms of selected Livonian nominals 207

Table 5.1: Finnish partitive and illative singular exponents 232

Table 5.2: Distribution of partitive exponents in Martin's (1995) data 235

Table 5.3: Genitive plural exponents in Finnish *s*-stem nominals 236

Table 5.4: The distribution of Finnish genitive plural exponents 237

Table 5.5: Uncertainly in the 'link' between lexical stem and inflexional suffix 239

Table 5.6: Inflexional patterns exhibited by old Occitan {PAIRE} 240

Table 5.7: Morphomic distribution of forms in Romanian feminine paradigms 241

Table 6.1: Estonian 'second' illative forms 256

Table 6.2: Discussion from §§6.2-6.4 consolidated 279

Table 6.3: Different loci in the languages discussed in this study 279

Table 7.1: Insensitive separative inflexional case-marking 284

Table 7.2: Insensitive cumulative and sensitive separative inflexional case-marking 284

Table 7.3: Languages in WALS with 6-7 case-forms 288

Table 7.4: Languages in WALS with 8-9 case-forms 289

Table 7.5: Formal variation across Wambaya inflexion classes in the genitive, proprietive, and origin cases 293

Table 7.6: The absolutive gender forms of the Wambaya private suffix 293

Table 7.7: Languages in WALS with 10+ case-forms 295

Table 7.8: Case suffix exponents in the Martuthuniran case system 298

Table 7.9: The distribution of Martuthuniran accusative and genitive suffixal exponents 298

Table 7.10: The distribution of Martuthuniran locative and effector suffixal

exponents 300

Table 7.11: Selkup non-possessive forms of {NOM} ‘God, heaven’ 303

Table 7.12: Sandhi in Selkup (where C1 and C2 belong to different ‘morphemes’) 304

Table 7.13: Morphophonemics in Selkup 305

Table 7.14: 1st singular possessive forms of Selkup {NOM} 306

Table 7.15: 2nd and 3rd singular possessive forms of Selkup {NOM} 308

Table 7.16: Latin third conjugation {REGŌ} ‘to rule, guide’ 311

Table 7.17: Continuants of Latin first person plural marker *-mus* 314

Table 7.18: Future forms in French {CHANTER} and Italian {CANTARE} ‘to sing’ 315

Table 7.19: The distribution of distinct stem allomorphs in Selkup 319

Table 7.20: Deduction of the genitive singular form of {QOŊ} ‘chief’ 321

Table 7.21: The paradigmatic distribution of distinct Latin verb stem allomorphs 323

LIST OF ABBREVIATIONS

1	First person
3	Third person
1Poss.	First person possessive
A.	Adjective
Abl.	Ablative
Acc.	Accusative
Com.	Comitative
cx	Case suffix
Dat.	Dative
El.	Elative
Fut.	Future
Fut. Perf.	Future Perfect
Gen.	Genitive
Gen.-Dat.	Genitive-dative
IA	'Item and Arrangement' model
Ill.	Illative
Impf.	Imperfect
In.	Inessive
Ind.	Indicative
Instr.	Instrumental
IP	'Item and Pattern' model
N.	Noun
No.	Number
Nom.	Nominative
Nom.-acc.	Nominative-accusative
Part.	Partitive
Perf.	Perfect
Pl.	Plural
Pluperf.	Pluperfect
Post.	Postposition
Prep.	Preposition
Pres.	Present
Sbjv.	Subjunctive
Sing.	Singular
SOV	Subject~Object~Verb word order
SVO	Subject~Verb~Object word order
Transl.	Translative
Translcom.	Translative-comitative
WP	'Word and Paradigm' model
V.	Verb

1.

INTRODUCTORY CONSIDERATIONS

1.0 THE TOPIC

The developments exhibited by inflexional case systems cross-linguistically have long been of interest to historical linguists. Three possibilities are found: languages can lose, maintain, or propagate morphological case distinctions through time. Latin is frequently cited as an example of the first type: in the modern Romance languages, only Romanian continues to distinguish case values morphologically (and then, only in feminine singulars) in nouns and adjectives. The robust five-value case system of Latin has thus almost entirely disappeared; many other Indo-European languages have exhibited similar developments.

The Finnic languages, on the other hand, present diachronic evidence of the maintenance and even accretion of inflexional case-marking. Finnish, for instance, still exhibits the large inventory of morphological cases inherited from proto-Finnic. In Veps, a closely related variety, innovative case-forms from postpositions and other postpositional material are in the process of grammaticalizing into morphological case-markers.

These facts have already been noted, and various explanations have been offered for the divergent developments exhibited by different varieties cross-linguistically. In reference to Latin and the Romance languages, phonetic erosion, analogy, and syntactic and functional considerations are amongst some of language-internal reasons suggested for the near-total loss of case in nouns and adjectives; external considerations, such as the result of contact with other languages, are also frequently cited as having had some effect on the development of the Latin case system.[1] In reference to case-increasing languages, scholars such as Korhonen (1996a) and Kabak (2006) have suggested that the presence of postpositions in a language facilitates the introduction of innovative inflexional suffixes. However, there are largely

[1] These endogenous and exogenous pressures will be considered in more detail in chapter two.

postpositional languages such as the Finnic variety Livonian which have lost inflexional case-forms through time. The reasons offered for the changes affecting case systems cross-linguistically therefore frequently account for the development of a specific language or a single case-form, but are insufficient to account more generally for why a language loses, maintains, or propagates inflexional case-marking.[2]

This study contends that, by directly comparing the ways in which the Italic and Finnic case systems have changed through time, explanations can be offered for the diachronic developments found in these, and other, languages. Close comparison of the developments found in a representative cross-section of these language branches reveals a major difference that correlates with the development of morphological case systems. *In nuce*: the morphological structures found in the inflexional case system and the ways in which these structures are exploited in language-use shed light on why certain languages have lost morphological case distinctions whilst others have maintained a 'steady' number of inflexional case-forms (or have even developed new morphological distinctions).

This study is conducted from an 'evolutionary' perspective of the type outlined by Croft (2000). It sees 'language' as a communication network, and language change as the result of complex systems of interacting 'pressures' which originate with the language-user. It therefore analyses language as a 'complex adaptive system' in which all aspects - phonology, morphology, syntax, semantics, pragmatics, and even social considerations, for instance - play complicated inter-related roles in determining what changes may come about in a given linguistic variety.

[2] Sapir's (1921) notion of 'drift' probably comes the closest to explaining the ways in which inflexional case systems develop cross-linguistically. However, as briefly pointed out by Croft (2000:4), "[i]f one speaks of 'forces' causing a language to change [...], then one is speaking as if language change is an inherent change applying to an abstract system", rather than recognizing that languages are changed through use by speakers, the latter view of which I take here and which is discussed in §1.1.0.

Before outlining the details of the 'evolutionary perspective' of language change in more detail, I will offer a brief note on some of the terms that will be used throughout this study. First, the term 'nominal' will refer *only* to nouns and adjectives. In respect of their inflexional morphology, nouns and adjectives in the Italic and Finnic languages tend to behave very similarly. Pronouns, demonstratives, numerals, and articles, on the other hand, often exhibit distinct morphological patterns (personal pronouns, for instance, frequently maintain a different number of case distinctions from nouns and adjectives cross-linguistically).[3] According to my use of the term, it is accurate to say that all Romance languages but Romanian lack a morphological case system in nominals, since it is in the pronominal system that morphological case is most faithfully maintained in modern Romance.

Second, the terms 'stable case', 'case-reducing', and 'case-increasing' have previously been used to describe the different development of case systems cross-linguistically.[4] Of these three, the latter two are adopted: 'case-reducing' languages are those which have unambiguously and overwhelmingly lost morphological case distinctions, whilst 'case-increasing' varieties are those which have unambiguously and overwhelmingly gained morphological case distinctions, through time. The term 'stable case', however, is a little more difficult to qualify as, synchronically, it is not always possible to determine how many case-forms to count in a given language,[5] and therefore it is not always clear whether the same number of formal distinctions has been maintained through time.[6] Moreover, in some languages (such as Finnish in §3.1), certain case values may have been lost whilst other forms have innovated. These varieties are therefore not unambiguously case-reducing nor unambiguously case-increasing.

[3] Iggesen (2011).
[4] Kulikov (2006; 2011).
[5] The problems associated with defining '(morphological) case' are analysed in §1.1.3.
[6] Amongst his 'stable case' systems, Kulikov (2011) includes those which 'reinforce' case-forms by means of the addition of an extra case marker or an auxiliary word and the 'borrowing' of inflexional patterns from other declensions. These systems therefore include those languages which resist erosion, not just those in which the same etymological forms are retained through time.

In order to side-step the problems associated with the term ‘stable case’, I will refer to those case systems which do not appear to have unambiguously increased or reduced throughout the attested history of the language in question as ‘case-steady’. This term is preferred here as it does not deny that case in these varieties is likely to have altered phonologically or functionally, and that the net number of case distinctions may have shifted one way or the other, but acknowledges that it is not clear that they should be uncontentiously analysed as ‘case-reducing’ or ‘case-increasing’. In reference to each of the languages explored in this study, I will offer explicit reasons for why they are analysed as ‘case-increasing’, ‘case-reducing’, or ‘case-steady’.

The structure of the remainder of this chapter is as follows: in §1.1.0 the ‘evolutionary’ perspective from which this study is conducted will be introduced in more detail; the central notions of ‘diachrony’, the ‘morphological pattern’, and ‘case’ will be considered from this point of view. In §1.2.0, the Italic and Finnic languages discussed in this study will be introduced. Finally, §1.3 will outline the content and structures of the following chapters.

1.1.0 AN EVOLUTIONARY PERSPECTIVE

An evolutionary perspective lends explanatory power to many of the developments that are found in the Italic and Finnic case systems. The desiderata underlying Croft’s (2000:4-6) evolutionary approach are that a theory of language change must:

1. avoid the reification or hypostatization of language;
2. offer explanations as to why languages do *not* exhibit particular developments, sometimes even over many generations of speakers;
3. distinguish two processes of change: the innovation (or actuation) and propagation (or diffusion), and/or loss of forms in a language;[7]

[7] Croft (2000:4) refers to innovation as ‘altered replication’ and propagation as ‘differential replication’.

4. subsume structural (e.g., syntactic), functional, and social dimensions (or their equivalents) of language change;
5. subsume both *internal* and *external* causes of language change.

Only a model of language change which meets all of these prerequisites may accurately account for changes occurring diachronically - given, that is, sufficient information about the phonology, morphology, syntactic structures, etc. of a linguistic variety, in addition to understanding about diastratic and diatopic variation, the influences of other languages with which it is in contact, and the situations in which different dialects and/or languages are used. The language-user is centralized in an evolutionary approach, and the role played by speakers and hearers, as well as their combined efforts to communicate and understand, provide insight into language development.

The evolutionary perspective is consistent with the view of language as a 'complex-adaptive system', a term developed in reference to study of living (i.e., biological) networks.[8] In reference to language, this view contrasts with the static systems encapsulated by prescriptive grammars, and involves the following central aspects:

a) there are multiple agents (consisting of speakers in the speech community);
b) the agents' behaviour is modelled on past interactions; both past knowledge and current use of language shapes the future behaviour of speakers;
c) the behaviour of a speaker is the result of various competing factors, including perceptual mechanics, the 'principle of least effort',[9] and social motivations;
d) the emergent structures by which a language is characterized are the result of "interrelated patterns of experience, social interaction, and cognitive processes."[10]

[8] As discussed by, e.g., Capra and Luisi (2014).
[9] Zipf (1949).
[10] Beckner et al. (2009:2).

According to the complex adaptive, or 'systems', view, it is the language-user that alters language, and who is therefore considered to be at the centre of the changes that occur cross-linguistically.

There are crucial differences between discussing evolution, for instance, and language as a complex adaptive system. In the first place, the basic 'agents' of change in evolution are: random cellular mutation, 'gene trading' (horizontal genetic transfer across bacteria), and "the creation of new forms of life through acquisition of entire genomes" or 'symbiogensis'.[11] The processes of propagation are, however, more restricted and predictable:

> "[w]hen new genetic patterns are created, they need to be integrated into their cellular environment. This involves a complex dynamic of an entire network of chemical reactions, in which only a limited number of new forms and functions are possible [...]. The entire process is far from random. Although there are random elements in the generation of new genetic patterns, from then on the process is highly ordered and complex."[12]

The propagation of changes in nature can be predicted with a great deal more certainty than the 'mutations' which occur in language. This is because, in language, the language-user is both the introducer and propagator of change, and the 'human' element in the use (and, by extension, the development) of a language is notoriously unpredictable. The social pressures surrounding the selection and propagation of variation are particularly difficult to forecast. For example, whilst instances of phonetic erosion might be expected, it cannot be predicted with certainty whether the change will occur uninterrupted, or whether, because the *lack* of phonetic erosion is associated with a 'desirable' speech community,[13] it will be 'reversed' so that speakers can maintain an association with a particular social group. The human element in language makes absolute predictions about what developments will be exhibited by a given

[11] Capra (2015:247).
[12] Capra (2015:247).
[13] The 'speech community' refers to the person or people with whom an individual shares experience and expertise, and with whom therefore they share common ground (Clark 1996) or mutual knowledge. (Discussed by Croft 2000:93f.).

variety through time impossible to offer. This study is therefore not concerned with offering precise predictions about what *will* happen, but with offering explanations about what *has* happened in the history of the Italic and Finnic languages. I will thereafter offer probabilistic generalizations about what might be *likely* to happen cross-linguistically in reference to the development of case systems (that is, whether it is likely that a variety will lose, maintain, or propagate inflexional case-marking), with an awareness that these suggestions are speculative at best.

The following sections will discuss in more detail the concepts of 'diachrony', 'morphological patterns', and 'case' from an evolutionary (or complex-adaptive system) perspective so that they may be used in a consistent way throughout this study.

1.1.1 DIACHRONY AND THE LANGUAGE-USER

As was briefly discussed in the previous section, according to an evolutionary perspective, change in language originates with the language-user. Language exists only by being used and perceived; it manifests in a network of speakers and hearers, writers and readers.[14] When it is used, there is the opportunity for change to occur since, in attempting to communicate with an audience, a speaker may adapt their vocabulary, their pronunciation, their syntactic constructions, etc., in order to better convey what they wish to express, or to more effectively establish themselves as part of a given speech community.[15] In this way, language varies not only diastratically or diatopically; vacillation of forms and structures can - and frequently do - occur within a single language-user's experience.

It is considered here that changes through time are the result of cumulative shifts in the mutually-accepted linguistic conventions employed by language-users within a given speech

[14] Croft (2000:90).
[15] Referring to the social psychology theory of 'accommodation' as discussed by Trudgill (1986) and Giles and Smith (1979).

community. That is: where a large enough number of people have all 'accepted' a particular form within their speech community which is distinct from that which may previously have been preferred, it constitutes an example of diachronic change. Changes may be introduced by language-users who may use, e.g., a new word in one environment, and who 'transfer' it into a second setting where it then gradually spreads through the speech community (and is perhaps introduced to yet others) and thus propagates. In this way, language change originates from language-use, and is not the result (only) of imperfect acquisition, of abstract 'drift', or of any sort of Greenbergian universals of language change.[16] Recognizing the language-user as central to language change does however necessitate the acknowledgement that that language change is difficult (if not impossible) to predict, given that many different factors will affect the development of a given language (or, rather, a small part of a given language). Nevertheless, throughout this study I will refer frequently to the language-user's role in language change; occasional mention will also be made of the acquisition process and what effect (if any) this may have in the loss, maintenance, or accretion of inflexional case-marking.

1.1.2 PATTERNS IN MORPHOLOGY

The most important evolutionary consideration for the sake of the current study is that of the vital and powerful human ability to recognize 'patterns' amongst various experiences, and to perceive 'meaningful' correlations between actions and reactions, or events and results, which may or may not be causal. Humans are by their very nature pattern-seeking - and pattern-imposing - creatures.[17] We tend to find or impose patterns that we deem 'informative' in all areas of our existence in order to more easily to make sense of the world around us; this

[16] Referring to the universal patterns of language identified by Greenberg (1963), wherein it is expected that the observed tendency should hold absolutely cross-linguistically. The approach adopted here is that there is no such thing as an absolute universal of language. An alternative perspective is discussed in §7.0.
[17] Kuhl (2000).

understanding is then utilized in survival. Language is no exception to this generalization, except that here the goal of pattern-recognition and -utilization is successful communication.

Let us take a simple example of a 'patterned' occurrence in human experience: that of the correlation between dark clouds in the sky and rainfall. Humans long ago learned to associate the two by means of previous contiguous occurrences of both dark clouds and rain, and can now take the former as the herald of the latter. This patterned occurrence of events (which in this instance happens to be causal: the clouds appear dark when they contain more moisture which allows less light to penetrate through them) enables us to better survive: by taking the gathering of dark clouds as a warning for rain, we may take shelter and avoid exposure, hypothermia, etc..[18]

This type of 'pattern' might not be the first that the reader considers when this term is used. Instead, the term 'pattern' may lead one to think of organized arrangements of dots and/or lines set up in aesthetically pleasing shapes and sequences. But from an evolutionary point of view, the human ability to recognize structures - whether physical or consequential - in the world around us, and to extend this recognition to novel circumstances, can be used to make familiar the unfamiliar and to reduce uncertainty about what may or may not happen in a given situation. To say, therefore, that the relationship of the dark clouds and bad weather constitutes a 'patterned' occurrence of events may be unusual, but, I contend, accurate.

Language in general and morphology in particular work in very similar ways to the example given above: in language use, we rely on identifying and utilizing different types of 'patterns' in order to aid communication. Although it is widely recognized that all humans seek and utilize patterns, and although this ability to recognize patterns is frequently

[18] This sort of situation bears many resemblances to the Pavlovian conditioning response as explained by, e.g., Rescorla (1988), as well as the Rescorla-Wagner model of conditioning (Rescorla and Wagner 1972; Wagner and Rescorla 1972). The Pavlovian conditioning response presents more evidence for the hypothesis that contiguous stimuli are recognized not just by humans but more generally in the animal kingdom.

discussed either implicitly or explicitly in reference to the study of morphology, it is not always clear what does or does not constitute a 'morphological pattern'. I will here discuss the way in which the term will be used in this study.

The definition offered by the Oxford English Dictionary for the term 'pattern' is certainly not adequate as a definition of a *morphological* pattern:[19]

> "A model, example or copy [...] [s]omething shaped or designed to serve as a model from which a thing is to be made; a design, an outline; an original [...] An example or model to be imitated; an example of particular excellence; a person or thing which is worthy of copying; an exemplar; an archetype."[20]

These definitions are telic in nature: the model is *designed* and is to be used *for the purpose of* copying; the example is *to be* imitated or *worthy of* imitating. A pattern here appears to be conceived of as something that is specifically created *so that it may be copied.*

In morphology, such implied telicity is not ideal. After all, patterns do not exist in language *in order to be* copied throughout the system. On the contrary, in morphology this telicity is somewhat flipped on its head: patterns are not created *to be* extended but instead, where a pattern can be detected by a language-user, it can be exploited in the inflexion of other lexemes; where formal variation - whether below, above, or at the level of the word - is not 'informative', or where it is opaque, the alternation will not be extended to other lexemes. The definition of a 'morphological pattern' must therefore avoid the same telicity as is implicit in the Oxford English Dictionary definitions of the term 'pattern'. I suggest the following better captures the ways in which the term (morphological) pattern will be utilized throughout:

---

[19] Nor, in fact, as a definition for a 'patterned occurrence ' such as was discussed above in reference to the recognition and utilization of correlated events in survival.

[20] http://www.oed.com/view/Entry/138977?rskey=CNDfWv&result=1&isAdvanced=false#eid. Accessed 20/02/2019, 14:25.

> **(Morphological) pattern:** "A formal alternation with which some systematic difference in another domain is perceived to correlate; the lack of formal variation where elsewhere a difference in content (whether semantic, functional, or morphological, for instance) is formally distinguished."

The rest of this section will be concerned with explaining this definition in more detail.

First, a 'formal alternation with which some systematic difference in another domain is perceived to correlate' is considered here to constitute a pattern only when it is extended to *at least one* other form in the language. Many patterns, therefore, will not spread beyond a handful of lexemes, and might only be restricted to creative language use. Nevertheless, the contention here is that, where a language-user extends formal variation beyond its original scope, it provides evidence that the alternation is one to which the language-user is sensitive.

Second, the 'other domain' may refer to phonological, semantic, functional, 'purely' morphological,[21] diatopic, or diastratic differences that are perceived to correlate with a formal alternation. A *functional difference*, for instance, pertains to distinctions in, e.g., number, case, person, tense; a *diatopic difference* refers to a difference which is associated with distinct geographical locations, etc.. In this study, I am more concerned with identifying and discussing *formal* alternations than *functional* ones,[22] and it must therefore be stressed that *purely morphological* information also falls under the purview of 'another domain'.[23] That is, in many languages, formal variation in the suffix provides information about the way in which the lexeme inflects.[24] In addition, therefore, to capturing differences in semantics, in

[21] In the sense of Aronoff (1994).
[22] Functional differences are often more difficult to describe and delimit than formal ones. The 'encyclopaedic approach' to linguistic meaning, according to which language-users 'build up' a semantic scope which is sketched out by all previous occurrences and uses of a particular form, is adopted here (Croft 2000:105-110). However, this means it is even more difficult to offer a definition of what constitutes a 'functional' or 'semantic' difference in a morphological pattern, particularly as it is likely to differ even between individuals.
[23] I am concerned only with the development of morphological case-*forms*. Where an inflexional distinction between 'nominative' and 'genitive' is lost, for instance, it does not mean that these functions disappear from the language, but I will not trace in detail the changing ways in which functions are expressed through time.
[24] This will be discussed further in §5.2 in reference to 'indexical' information provided by morphological variation.

grammatical function, in register, style, or dialect, formal variation may also be associated with information about declension class membership or other inflexional behaviours.

The term ‘domain’ may also refer to the purely phonological, omitting semantics from consideration. In many examples from folk etymology, for instance, phonological similarity is sufficient for the recognition and extension of patterns which involve the partial or total replacement of opaque elements in wordforms with more transparent, though often semantically ‘nonsensical’, sub-word parts.[25] Whilst the definition above is intended to capture this aspect of pattern recognition, it will not be considered in more detail in this study.

The third consideration to be noted in reference to the definition above is that, for a pattern to be recognized, some formal similarity between the two (or more) forms which are to take part in the morphological pattern is required. The phonological distinction between English /kæt/ and /dɒg/, for instance, correlates with a semantic difference ‘feline’ and ‘canine’, which could conceivably be extended (albeit to a limited extent) to differentiate other members of the feline and canine animal groups. However, since the formal variation is so drastically different (the two phonological forms have no sounds in common), we do not find the extension of the semantic distinction according to morphological differences exhibited by the lexemes {CAT} and {DOG}. Moreover, in most instances it is not sufficient only that there is a phonological alternation between two or more forms: for the formal variation to constitute a pattern that may be extended there must (usually - cf. the discussion on folk etymology above) also be a contiguous functional (morphological, etc.) correlate which a language-user can associate with the difference between [FORM1] and [FORM2]. The formal difference between English /mæp/ and /næp/, for instance, consists only of the voicing of the initial nasal - the two forms are thus phonologically very similar. However, there is no

[25] Maiden (2007).

systematic functional distinction between the referents of these two nouns which can be extended elsewhere in the language. On the other hand, the morphological variation that occurs between the forms /kæt/ and /kæts/ correlates with a semantic alternation that is much more widely utilized - that of 'singular' versus 'plural'. The phonological similarities outweigh the differences, and there is a transparent morphological variation which can be recognized, particularly when other forms with a similar phonological alternation (and functional alternation) such as /mæt/~/mæts/ and /ræt/~/ræts/ are taken into consideration.

Finally, it is not just formal variation that may be extended across the language. In fact, morphological *invariance* is frequently recognized and extended to novel forms.[26] One example is that of the lack of formal distinction of nominative and accusative functions in the plural of Latin third declension nominals, which, it has been hypothesized, has been extended to first declension lexemes in certain inscriptions from Africa and Gaul.[27] In this example, it appears as though a *lack* of morphological variation has been extended to lexemes belonging to a declension class in which nominative and accusative case values were once formally differentiated in the plural. The syncretism of the Latin third declension nominative and accusative in the plural, I argue, thus constitutes an example of a morphological pattern, despite not being an instance in which formal *variation* is extended to novel forms.

The definition offered here is a preliminary one. It does not allow one to identify the 'limits' of morphological patterns (i.e., to determine what does or does not constitute a 'morphological pattern'), nor predict what types of formal alternation may become associated with functional (morphological, etc.) distinctions (and therefore potentially become 'productive'). In the first place, I am myself not sure that there *are* limits as to what can constitute a 'morphological pattern', so long as the criteria given in the definition above are

[26] N.B., throughout, 'novel forms' refer to previously unencountered inflected forms rather than neologisms after Blevins (2016).
[27] This development is discussed in §2.1.3 and §2.1.5.

met.[28] In chapters two and three in reference to the Italic and Finnic languages respectively, for instance, morphological patterns will be identified at sub-suffixal- (e.g., the Finnish excessive, §3.1.3) up to the word-level (e.g., Romanian case/number alternations in nominal paradigms, §2.3). Evidence will be presented (in chapters two, three, and five in particular) which demonstrates that language-users recognize formal variation at these levels and associate it with particular functional (morphological, etc.) differences.[29] I therefore suggest that, as long as formal variation (or invariance) which occurs in a language may be successfully associated by at least one language-user with some difference in another domain and extended to a novel form in at least one instance, then any formal alternation (or lack thereof) existing between two or more forms constitutes a morphological pattern. For the purposes of this study, therefore, the definition above will suffice.

1.1.3 A LANGUAGE-USER PERSPECTIVE ON 'CASE'

Any study of 'case' faces issues from the very outset in that there is no universally satisfactory or accepted definition of what a 'case' actually is. Throughout literature which discusses the concept, the term is not always used consistently, and scholars infrequently offer explicit elucidation about how they employ it in their work, which can lead to ambiguity. The term is often used to refer to as many as three distinct notions:

- the *form* of a given lexeme if it is distinct from another in the paradigm;
- the *name* traditionally given to forms which may differ across declensions but which nevertheless have the same syntactic distribution. For instance, even though

[28] Since I am concerned with all and any patterns than provide information in the production of novel forms, I do not consider it essential here to differentiate those patterns which are likely to generalize from those which are not. cf. Blevins (2009) and references therein.

[29] Although this study will not be concerned with constituents above the level of the word, it is very likely that variation exhibited by larger units such as periphrastic constructions and idioms may also offer examples of 'morphological patterns'.

they are phonologically distinct, the Latin forms MENSĀ, EXERCITŪ, and RĒGE can all be said to be 'in the ablative (singular) case';[30]

- the *function(s)* a form encodes. For example, an inflected nominal ending in *-lle* in Finnish encodes 'allative' functions.

In order to avoid ambiguity, I will use different terminology to distinguish each of these notions. Throughout this study, I use the term '[case] form' (for example, the 'nominative (singular) form', 'genitive (plural) form', etc., although without inverted commas) to refer to the phonological shape of a fully-inflected lexeme; the different names of case values will be used to refer to inflected wordforms which encode the same functions regardless of differences in morphological shape (for instance, the Latin forms MENSĀ, EXERCITŪ, and RĒGE can be described as all inflected in the 'ablative [singular] case'); finally, the functions expressed by distinct case-forms will be given in inverted commas, e.g., 'patient', 'allative', 'away from' etc..

The bare term 'case' will be used in a more general way to refer to the "system of marking dependent nouns for the type of relationship they bear to their heads".[31] Although this is a commonly-cited definition of case, it is frequently acknowledged to be in some ways insufficient. First, it does not differentiate 'morphological' ('inflexional', 'synthetic') from 'syntactic' ('analytic') marking. Case values can be distinguished inflexionally: by an affix; by a morphophonological difference in the shape of the lexeme (with or without the addition of overt affixal material); or by alternations in stress or pitch. A wordform can also express particular syntactic relations simply by virtue of its place in the inflexional paradigm. The syntactic distinguishing of case includes adpositional and clitic marking, as well as constituent order. Cross-linguistically, both morphological and syntactic case-marking can

[30] Throughout, length will be marked on Latin lexemes where it is relevant to the morphological analysis.
[31] Blake (1994:1).

differentiate the same sorts of functions (e.g., 'agent', 'patient', 'source', 'goal'), and both are found in all of the Italic and Finnic varieties discussed here.

In this study, the notion of 'case' is approached from two different perspectives. In the first place, it is in my interest to separate synthetic and analytic marking, since I am concerned only with the way in which *inflexional* case systems develop through time. Only by offering means of differentiating these types of case-marking is it possible to delimit inflexional case and comment on the changes which have affected these forms diachronically. In the second place, from an evolutionary perspective, the similarities exhibited by morphological and syntactic case are essential for understanding certain changes found in the languages under consideration. The language-user, who drives change in language, appears to be sensitive to similarities exhibited by the behaviour of synthetic and analytic case (as demonstrated by the frequent instances of the grammaticalization of postpositions into innovative case suffixes cross-linguistically).[32] Both the differences and the similarities between affixes and adpositions must therefore be addressed in this section.

Distinguishing the morphological from the syntactic expression of case is not always straight-forward, as case-forms in real language do not always behave strictly as one or the other. In order to differentiate morphological case-forms and adpositions, whilst also allowing for recognition that these forms in natural languages do not always accord with the prescriptive definitions given for them, I approach their description from a canonical perspective (e.g., Corbett 2007 and 2009). The canonical approach in linguistics is one that delimits what is *possible* in order to situate what is *actual*, allowing the linguist in this instance to offer canonical definitions for both 'morphological' and 'syntactic' case in order to identify the logical 'endpoints' of the two, and to more accurately differentiate

[32] Blake (1994:166-168); Kulikov (2011:440-444).

morphological and syntactic case (as well as recognize the ways in which they are similar) in the Italic and Finnic languages.

The canonical morphological case is the inflexional expression of syntactic relations that, for example, a noun bears to other constituents in a sentence. It will obey language-specific phonological constraints (such as vowel harmony), and licence concord; it cannot be separated from the nominal with which it occurs, shows no degree of syntactic autonomy (i.e., autonomy beyond the level of the lexeme), and cannot be 'interrupted' by other words (although inflexional material may occur between a nominal stem and an inflexional case suffix). The canonical adposition, on the other hand, behaves as a separate phonological word (e.g., can take main stress), does not obey language-specific phonological constraints (such as vowel harmony), and is formally invariant for, e.g., inflexion class. It is possible for whole words to intervene between a governed nominal and its adpositional head; the adposition occurs only once in a phrase.

Finally, inflexional case tends to encode more 'grammatical' relations than adpositions, which in their turn tend to encode more 'semantic' functions.[33] For instance, in Latin, the expression of 'subject' and 'direct object' is restricted to inflexion; for meanings such as 'on top of' or 'underneath', prepositions such as SUPER and SUB respectively are used. This diagnostic tool is, however, usually insufficient to distinguish synthetic and analytic case, as the difference between what constitutes a 'grammatical' and a 'semantic' meaning is often difficult (if not impossible) to determine.

Throughout this study, the term 'case' refers exclusively to the inflexional means of expressing syntactic and semantic information; the term 'adposition' will be used for the syntactic means of expressing these functions. Furthermore, although throughout I will refer

[33] Blake (1994:11), Hagège (2010:38), and Kittilä, Västi, and Ylikoski (2011:4).

to the ‘case systems’ of the Italic and Finnic languages, it should be understood that this refers to the *inflexional* or *morphological* case systems of these varieties, and excludes all other means of expressing ‘case’ functions (e.g., word order, adpositions, clitics).

It is unlikely that the language-user has available strict definitions for morphological and syntactic case-markers, although there may be an intuitive sense of the ways in which they behave differently. (For instance, even those with no linguistic training may be sensitive to the fact that the adposition and the nominal with which it occurs can be ‘interrupted’ by other lexical material, whilst the affix and the nominal cannot.) That they are also sensitive to the *similarities* between inflexional affixes and adpositions is demonstrated by the diachronic data, particularly in the relatively frequent grammaticalization of postpositions into morphological suffixes (discussed in reference to Veps in §6.4).

The canonical approach allows us to differentiate morphological and syntactic case, and thus discuss only the former. The language-user perspective, on the other hand, requires that we recognize the behavioural similarities of inflexional affixes and adpositions which facilitate the grammaticalization of the latter into the former diachronically. It is therefore inevitable, from an evolutionary perspective in which the language-user as an originator of change is centralized, that both morphological and syntactic case will be discussed in this study. The canonical view, however, provides a point from which to start when it comes, in the first place, to determining how many inflexional case-forms are found in a given language, and, in the second place, to tracing the development of novel suffixes as they grammaticalize from syntactic into morphological markers of case.

1.2.0 THE DATA SET

It is not possible here to offer in-depth descriptions of the development of case in all Italic or Finnic varieties. Instead, a representative sample which encapsulates the relevant

synchronic and diachronic aspects of the case systems found in these languages has been selected. In the following sections, I offer a brief introduction to these Italic and Finnic varieties.

1.2.1 THE ITALIC LANGUAGES

Throughout this study, Latin and the Romance languages will be collectively referred to as the 'Italic' languages. Discussing just the 'Latin' case system or the case system of the 'Latin languages' would introduce ambiguity, as it would not be clear whether this refers to the case system of Latin or the case systems of Latin *and* Romanian, for instance. The term 'Italic' does *not* here refer to the long extinct languages in Italy which were spoken contemporaneously with Latin, such as Oscan and Umbrian.[34]

In this study I focus on Latin (§2.1), old French and old Occitan (§2.2), and Romanian (§2.3). The term 'old Gallo-Romance' will be used to refer to old French and old Occitan collectively;[35] where old Gallo-Romance and Romanian are discussed separately from Latin, they will be referred to as the 'Romance' varieties. The morphological case system of Latin is found in the earliest attested stages of the language, but has been lost through time; the vast majority of modern Romance varieties do not distinguish case morphologically in nouns or adjectives. Romanian and old Gallo-Romance are the only Romance languages which maintain (very limited) case systems; these varieties are therefore included in the discussion here.

A brief discussion of what is meant by the terms 'Latin', 'old French', 'old Occitan' and 'Romanian' must be offered. 'Latin' is a useful but in some ways inadequate name for a

[34] Consideration of the development of inflexional case in these languages is left to future research.
[35] Other old Gallo-Romance varieties will not be discussed here on consideration of space, and because many of the observations offered in reference to old French and old Occitan also apply to other old Gallo-Romance varieties.

language variety that is first attested from around the seventh century B.C.E.;[36] it was initially spoken in central Italy and later spread with the expansion of the Roman Empire. The term 'Latin' suggests an homogeneous entity with a single grammar persisting through time (as does any language name, in fact). However, despite the apparent lack of diachronic, diastratic, and diatopic variation in the written variety, there is evidence to suggest that, much like other, living languages, there was a great deal of social and regional variation at all stages of the Latin language.[37]

A further complicating factor in delimiting and defining Latin is the fact that there is no easily-identified 'end-point' of Latin and 'starting-point' of the Romance languages, and it is therefore in some ways difficult to offer a definition of one that excludes the other. Although much has been written on what date Latin 'became' the Romance languages,[38] it is inaccurate to talk of the 'end' of Latin, since the language has been continuously transmitted from parent to child since its earliest attestations. There is therefore no neat line dividing 'Latin' from 'Romance', and any definition offered must be given with the understanding that some of the aspects characteristic of the Latin language may also be found in some or all Romance varieties.

This study need not be concerned with determining where 'Latin' ends and 'Romance' begins, only with offering some means of differentiating 'Latin' from 'old Gallo-Romance' and 'Romanian'. Thus, the term 'Latin' here refers to the linguistic variety for which we have only written evidence and which was spoken, at the height of the Empire, across "much of western and central Europe, the Mediterranean, and along the coast of North Africa",[39] and which was, importantly, characterized by more than three morphological case-forms. This

[36] Penney (2011:220).
[37] Some of the evidence for this variation will be consulted in chapter two.
[38] For instance, Lot (1931), Clackson and Horrocks (2007:265-304), and Wright (2011:59-79) and (2013).
[39] Ledgeway (2012:1).

definition therefore includes within its scope the postulated 'vulgar' Latin (a term that will be avoided in this study) which maintained a reduced case system when compared to the 'Classical' language, but excludes those Romance varieties which exhibit a two-value case system.

The term 'old French' will be used to refer to the linguistic variety that was used in the north of Gaul; 'old Occitan' was spoken in the southern regions. For the purposes of this study, these old Gallo-Romance varieties are distinguished from Latin (as well as from the modern varieties French and Occitan) by being characterized by two case values. I opt for this criterion, rather than offering dates for when these varieties were used, because there is generally no firm consensus on the exact dates between which the 'old French' and 'old Occitan' varieties were spoken.[40] I am only concerned with those texts from the times and places where there was still some indication that inflexional variation could be associated with a functional (stylistic, etc.) difference.

Finally, 'Romanian' is the name given to the official language of the Republic of Romania (and the Republic of Moldova) and is currently spoken "over a territory which roughly includes the Roman province of Dacia, but also areas historically outside the Roman Empire", including northeastern Bulgaria, Serbia, Hungary, and Ukraine.[41] Scholars generally identify two diachronic 'varieties' of Romanian: old Romanian and (modern) Romanian. Old Romanian texts date from between 1500 and 1780;[42] texts from after the end of the eighteenth century until the modern day are generally considered to be written in 'modern' Romanian. For the most part in this study I will focus on inflexional variation found in the modern language. Moreover, since the two periods do not exhibit many differences with

[40] Nyrop (1935), Pope (1952), and Price (1971) date 'old' French from roughly the middle of the ninth century to the beginning of the fourteenth; Paden (1998) and Skårup (1997) date 'old' Occitan from the eleventh to the thirteenth century.
[41] Maiden (2016a:91).
[42] Martin Maiden (p.c.); cf. Chivu et al. (1997) and Hill and Alboiu (2016:xiv).

regard to the inflexional case systems (i.e., in the number of case distinctions found or in the means of encoding these distinctions), the need of differentiating them is less important than differentiating, for instance, 'old French' and 'modern French'. I will therefore use the term 'Romanian' to refer collectively to both the old and modern varieties.

1.2.2 THE FINNIC LANGUAGES

For the most part, the Finnic data are less well-known than the Italic. Generally, therefore, the varieties discussed here will be described in greater depth than Latin, old Gallo-Romance, and Romanian.

Finnic belongs to the Uralic family and includes two 'major' varieties, Finnish and Estonian, as well as a number of smaller languages. The living Finnic varieties are largely spoken in and around Finland, Estonia, and western Russia, and share numerous lexical, syntactic, and morphological similarities which are particularly clear in the case system. All Finnic varieties maintain case systems with eight or more morphological case-forms, which most traditional grammars split into 'grammatical' case-forms (those which encode syntactic relations such as 'subject', 'agent', and 'patient', i.e., the nominative, genitive, and partitive; the dative in Courland Livonian), and the 'semantic' case-forms, which largely express more peripheral meanings such as 'source', 'goal', and 'abessive' functions (the remaining case values in the Finnic languages). Most nominals inflect for two numbers (singular and plural) in addition to case; no Finnic variety exhibits grammatical gender.

Unlike the Romance data, there is no extant evidence of the earlier stage of the Finnic languages from which the modern varieties have developed. Reconstructions of 'proto-Finnic', its case system, and developments that have occurred during the unattested history of

these varieties are therefore necessary, but are, of course, speculative in nature.[43] Nevertheless, occasionally I will refer to proto-Finnic and, where I do so, I will offer brief reasons for why the allusion is necessary and explain where reconstructions are problematic.

In chapter three, I will describe the case systems of three Finnic varieties: Finnish, Livonian, and Veps. Finnish (§3.1) has been chosen for this study because it is considered the 'most conservative' Finnic variety,[44] exhibiting minimal phonological erosion and many aspects reconstructed for proto-Finnic (such as vowel harmony).[45] In addition, Finnish is the largest Finnic variety and has been the object of extensive study, which makes it an accessible source of data. Livonian and Veps are two smaller Finnic dialects rarely discussed outside of literature specifically on the Finnic and Uralic languages. Livonian (§3.2) was natively spoken in the north of Latvia until recent years. It maintains the smallest morphological case system of any Finnic language, and there has been a great deal of phonetic erosion since its split from the proto-language. Veps (§3.3) is currently spoken in certain areas of Russia by some three and a half thousand people.[46] In this variety there are a number of innovative morphological case-forms that have developed (and are continuing to develop) from postpositional material since the splitting up of proto-Finnic. It maintains the largest morphological case system of any Finnic variety.[47]

The developments exhibited by Finnish and Veps represent the changes that affect all other Finnic languages but Livonian - that is, the maintenance or accretion of morphological

[43] In fact, the nature of the written evidence of the Italic languages means that observations about the development of the Latin case system are also frequently speculative.
[44] Grünthal (2003:24 *et passim*); cf. Austerlitz (1993:29) and Abondolo (2006c:149).
[45] Assuming that Finnish retains elements found in proto-Finnic and reconstructing proto-Finnic aspects based largely on evidence from Finnish is one of those instances in which arguments about the earlier stage of these languages is in danger of being circular.
[46] Puura et al. (2013:4).
[47] Where the case systems of Estonian and of the other, smaller Finnic dialects exhibit relevant differences or similarities to Finnish, Livonian and Veps, they will be discussed.

case-forms since the splitting of proto-Finnic.[48] Livonian is the only case-reducing Finnic language. In this way, the selection of these specific varieties offers an exhaustive representation of the types of developments that are found in this linguistic branch of the Uralic family.

A brief note is needed on terminology. All of the Finnic languages discussed here are characterized by an abundance of local case-forms which are usually split into 'internal' and 'external' types, depending on the functions they encode; in Veps a third 'approximative' local case series is also identified. These are further split into three directional subtypes: 'source', 'goal', and 'stasis':

	External	**Internal**	**Approximative**
Source	'away from' (ablative)	'from inside' (elative)	'away from the vicinity' (egressive)
Stasis	'on' (adessive)	'in' (inessive)	'in the vicinity' (approximative)
Goal	'towards' (allative)	'into' (illative)	'towards the vicinity' (propinquative)

Table 1.1: General functions of the Finnic local cases

Throughout, I will refer to, e.g., the ablative, elative, and egressive as the 'source' cases and to the, e.g., elative, inessive, and illative as the 'internal local case series', etc., following convention.

Finally, it should here be noted that the way in which inflexion classes in the Finnic and Italic languages are identified differs. That is: where declension classes are distinguished in Latin, old Gallo-Romance, and, in some accounts (§3.3.1), Romanian by suffixal variation, since the vast majority of suffixal case markers particularly in Finnish and Veps are invariant

[48] Five values are usually reconstructed for proto-Uralic (Korhonen 1996b:222 and Abondolo 2006b:18-21), which means that all Finnic varieties have increased the number of case values distinguished morphologically through time. I consider here only the development of the Finnic case systems since the splitting of proto-Finnic, since it is since this time that the Finnic case systems exhibit distinct developments.

across number and lexeme, inflexion classes are distinguished instead by means of alternations in the lexical stem. Latin neatly exemplifies the first inflexion class type:

	Inflexion Class 1		**Inflexion Class 2**	
	Singular	**Plural**	**Singular**	**Plural**
Nominative	AGRICOLA ‘farmer’	AGRICOLAE	DOMINUS ‘master’	DOMINĪ
Accusative	AGRICOLAM	AGRICOLĀS	DOMINUM	DOMINŌS
Genitive	AGRICOLAE	AGRICOLĀRUM	DOMINĪ	DOMINŌRUM
Dative	AGRICOLAE	AGRICOLĪS	DOMINŌ	DOMINĪS
Ablative	AGRICOLĀ	AGRICOLĪS	DOMINŌ	DOMINĪS

Table 1.2: Inflexion classes in Latin

In these examples, the lexical stem remains invariant within the paradigm. Across the two paradigms given in table 1.2, however, the shape of suffix differs in all cells but the dative plural and ablative plural. The formal variation of the suffix is how different declension classes are identified not only in Latin, but also in old Gallo-Romance and Romanian (as well as more generally across the Indo-European family). This compares with the stem-alternation means of distinguishing inflexion classes, as encapsulated by Finnish:

	Inflexion Class 1		**Inflexion Class 2**	
	Singular	**Plural**	**Singular**	**Plural**
Nominative	*vastaus* ‘answer’	*vastaukset*	*vieras* ‘stranger’	*vieraat*
Genitive	*vastauksen*	*vastauksien*; *vastausten*	*vieraan*	*vieraiden*; *vieraitten*
Partitive	*vastausta*	*vastauksia*	*vierasta*	*vieraita*
Essive	*vastauksena*	*vastauksina*	*vieraana*	*vieraina*
Inessive	*vastauksessa*	*vastauksissa*	*vieraassa*	*vieraissa*
Adessive	*vastauksella*	*vastauksilla*	*vieraalla*	*vierailla*

Table 1.3: Inflexion classes in Finnish

In these examples, the ‘link’ (§5.3.2) between the lexical stem and the inflexional material exhibits morphophonological alternations. Different lexemes are characterized by slightly different intra-paradigmatic formal variation, and it according to this that distinct inflexion classes are identified. Similar observations hold for Veps. Finnic declension classes will be described in greater detail in chapter three.

1.3 CLOSING REMARKS

This chapter has introduced the topic with which the rest of the study is concerned and has offered a background against which the work will be conducted. It has aimed to offer some elucidation about the ways in which particular terms will be employed in order that they are unambiguous in their usage.

The rest of this study is structured as follows: in chapter two, the case systems of Latin, Romanian, and the old Gallo-Romance varieties old French and old Occitan will be introduced in detail. Synchronic descriptions of each of these case systems will be given first, followed by an account of their developments. A similar overview of the Finnish, Livonian, and Veps case systems will be offered in chapter three, although a little more detail will be offered here as it will be assumed that the reader is not as well-acquainted with these languages as with the Italic varieties under discussion.

Chapter four will present the Italic and Finnic data from constructivist and abstractivist approaches, and assess the merits of each of these perspectives in reference to the case systems described in chapters two and three. It will be concluded that the abstractivist approach more adequately captures the ways in which novel forms are interpreted and produced in each of the languages considered here, but that the constructivist models highlight a key typological morphological difference between the two families. That is: the Italic languages are characterized by a great deal more ‘fusional’ inflexional structure in which form and function are rarely, if ever, in a biunique mapping at the sub-word level, whereas Finnish and Veps in particular exhibit predominantly ‘agglutinative’ inflexional

structure,[49] in which suffixal material can frequently be consistently associated with particular functions.

Chapter five will offer further elucidation of the two types of morphological structure identified in chapter four, introducing the terms 'substance entities' and 'schematic patterns' to describe them. This chapter will offer comparative diachronic evidence which indicates that speakers recognize and exploit both of these types of morphological structure in language use. Chapter six will then draw together the analyses of chapters four and five and explain how the different types of morphological structure found in the data considered help us better understand the ways in which case systems are likely to develop through time. It will be shown that clarity in sub-word structure and an abundance of instances in which form and function (or sets of functions, such as 'genitive' *and* 'singular') are in an isomorphic relationship is more conducive to the maintenance and even propagation of case systems, whilst frequent opacification of the form~function mapping at the sub-word level is more likely to lead through time to the loss of morphological case distinctions.

Finally, in chapter seven, the conclusions offered in chapter six in relation to the Italic and Finnic languages will be supplemented by consideration of the (synchronic) relationship between the morphological structure(s) found in language and the size of the case system with a larger data sample, mostly drawn from the *World Atlas of Linguistic Structures* online. Problems associated with the analyses offered at the end of this study will also be addressed in this chapter, particularly in reference to the differences exhibited by the development of the Latin verb system, as well as possible solutions and avenues for further research. Concluding remarks will be offered in chapter eight.

[49] Throughout, I will discuss agglutinative and fusional *structures*, as opposed to *languages*, since there are very few - if indeed there are any - examples of languages that are wholly agglutinative or wholly fusional. Here, the term 'agglutinative' refers to morphological structures in which function (or a set of functions) and form are in a perfect one-to-one mapping; the term 'fusional' refers to morphological structures in which the form~function mapping is interrupted in various ways (discussed in more detail by Anderson 2015).

2.

THE ITALIC LANGUAGES

2.0 THE ITALIC CASE SYSTEM

This chapter will focus on presenting the case systems of Latin (§2.1), old French, old Occitan (§2.2), and (modern) Romanian (§2.3) as they are traditionally given in grammars. Using these as points of reference, a sketch of some of the major phonological, analogical, syntactic, and contact-induced developments which have occurred in these case systems through time will also be offered. It will be demonstrated that no single type of development could have caused the near-total loss of the Latin case system, but that the (attested or hypothesized) changes that have affected the Italic case system are multifaceted and inter-dependent. Moreover, it will be concluded that, though an evolutionary view of the data allows for a more complete survey of the diachronic changes exhibited by the case system and particular case-forms, it does not adequately explain the general tendency exhibited by the Italic languages to lose morphological case-markers.

2.1.0 INTRODUCTION TO LATIN

In the following sections, I will outline the functions and forms of the (Classical) Latin cases (§2.1.1) as well as certain aspects of their diachronic development (§§2.1.2-5). I do not presume here to offer an exhaustive account of the Latin case system, but instead present those data which will be relevant to later discussion, with an awareness that much of what follows may be familiar to the reader.

2.1.1 THE LATIN CASE SYSTEM

Latin grammars generally list six case values: nominative, vocative, accusative, genitive, dative, and ablative; three gender values: masculine, feminine, and neuter; and two number values: singular and plural. Morphological case is distinguished by suffixes which also

cumulatively express number (and, in adjectives, gender). Most nominals belong to one of five major declensions:

	First Declension		Second Declension	
	Singular	**Plural**	**Singular**	**Plural**
Nominative	PUELLA ‘girl’	PUELLAE	SERUUS ‘slave’	SERUĪ
Vocative	PUELLA	PUELLAE	SERUE	SERUĪ
Accusative	PUELLAM	PUELLĀS	SERUUM	SERUŌS
Genitive	PUELLAE	PUELLĀRUM	SERUĪ	SERUŌRUM
Dative	PUELLAE	PUELLĪS	SERUŌ	SERUĪS
Ablative	PUELLĀ	PUELLĪS	SERUŌ	SERUĪS
	Third Declension - Masculine		**Third Declension - Neuter**	
	Singular	**Plural**	**Singular**	**Plural**
Nominative	RĒX ‘king’	RĒGĒS	OPUS ‘work’	OPERA
Vocative	RĒX	RĒGĒS	OPUS	OPERA
Accusative	RĒGEM	RĒGĒS	OPUS	OPERA
Genitive	RĒGIS	RĒGUM	OPERIS	OPERUM
Dative	RĒGĪ	RĒGIBUS	OPERĪ	OPERIBUS
Ablative	RĒGE	RĒGIBUS	OPERE	OPERIBUS
	Fourth Declension		**Fifth Declension**	
	Singular	**Plural**	**Singular**	**Plural**
Nominative	GRADUS ‘step’	GRADŪS	RĒS ‘thing, matter’	RĒS
Vocative	GRADUS	GRADŪS	RĒS	RĒS
Accusative	GRADUM	GRADŪS	REM	RĒS
Genitive	GRADŪS	GRADUUM	REĪ	RĒRUM
Dative	GRADŪI	GRADIBUS	REĪ	RĒBUS
Ablative	GRADŪ	GRADIBUS	RĒ	RĒBUS

Table 2.1: The major declensions in Latin

First and fifth declension nouns are largely feminine with no neuter members; second declension nouns are mainly masculine and neuter; the third and fourth declensions include members of all three genders. Neuter nominals inflect in a slightly different way to non-neuter nominals, where the accusative and nominative functions are not distinguished in the singular or in the plural (although the forms are distinct across number).

In addition to these ‘major’ declensions, there are a handful of smaller inflexion classes. For instance, those for loanwords from Ancient Greek:

	First Declension		Second Declension	
	Singular	**Plural**	**Singular**	**Plural**
Nominative	NYMPHĒ ‘nymph; bride’[1]	NYMPHAE	MYTHOS ‘myth’	MYTHĪ
Vocative	NYMPHĒ	NYMPHAE	MYTHE	MYTHĪ
Accusative	NYMPHĒN	NYMPHĀS	MYTHON	MYTHŌS
Genitive	NYMPHĒS	NYMPHĀRUM	MYTHĪ	MYTHŌRUM
Dative	NYMPHAE	NYMPHĪS	MYTHŌ	MYTHĪS
Ablative	NYMPHĒ	NYMPHĪS	MYTHŌ	MYTHĪS

Table 2.2: First and second declension Ancient Greek loanwords

Nominals belonging to these minor declensions exhibit a number of inflexional differences to those of native first and second declension Latin lexemes, such as the first declension genitive singular in *-ēs* (the Ancient Greek first declension genitive singular marker), and second declension accusative singular in *-n* (the Ancient Greek accusative singular exponent).

Finally, the majority of adjectives follow the inflexional patterns exhibited by first, second, and third declension nouns.

	First/Second Declension BELLUS, -A, -UM ‘pretty’					
	Masculine		Feminine		Neuter	
	Singular	**Plural**	**Singular**	**Plural**	**Singular**	**Plural**
Nominative	BELLUS	BELLĪ	BELLA	BELLAE	BELLUM	BELLA
Vocative	BELLE	BELLĪ	BELLA	BELLAE	BELLUM	BELLA
Accusative	BELLUM	BELLŌS	BELLAM	BELLĀS	BELLUM	BELLA
Genitive	BELLĪ	BELLŌRUM	BELLAE	BELLĀRUM	BELLĪ	BELLŌRUM
Dative	BELLŌ	BELLĪS	BELLAE	BELLĪS	BELLŌ	BELLĪS
Ablative	BELLŌ	BELLĪS	BELLĀ	BELLĪS	BELLŌ	BELLĪS

	Third Declension TRISTIS, -E ‘sad’			
	Masculine/Feminine		Neuter	
	Singular	**Plural**	**Singular**	**Plural**
Nominative	TRĪSTIS	TRĪSTĒS	TRĪSTE	TRĪSTIA
Vocative	TRĪSTIS	TRĪSTĒS	TRĪSTE	TRĪSTIA
Accusative	TRĪSTEM	TRĪSTĒS/TRĪSTĪS	TRĪSTE	TRĪSTIA
Genitive	TRĪSTIS	TRĪSTIUM	TRĪSTIS	TRĪSTIUM
Dative	TRĪSTĪ	TRĪSTIBUS	TRĪSTĪ	TRĪSTIBUS
Ablative	TRĪSTĪ	TRĪSTIBUS	TRĪSTĪ	TRĪSTIBUS

Table 2.3: Major adjective declensions in Latin

[1] There is also a nativized lexeme {NYMPHA} with the same meaning which declines like {PUELLA} in table 2.1.

Adjectives belonging to the first and second declensions inflect for all three genders in both singular and plural, whilst those belonging to the third declension exhibit no morphological differences between masculine and feminine in either number.

The main functions of the Latin cases are as follows:[2]

- nominative: the form of the subject of the verb;
- accusative: the form of the direct object of the verb and the complement of certain prepositions; the form of the subject in subordinate clauses in the 'accusative and infinitive' construction;
- genitive: most commonly used to qualify nouns; encodes possession and partitive functions (e.g., MULTĪ MILITUM 'many of the soldiers'); also modifies certain verbs and adjectives;
- dative: most frequently encodes the indirect object, i.e., the beneficiary or recipient; the form of the possessor in certain constructions with the verb {ESSE} 'to be'; also used to modify certain verbs and adjectives;
- ablative: the bare ablative form can encode three main functions: 'in, on, at' ('locative'), 'from' ('ablative'), and 'with, by' ('instrument/agent'); also used as the form of the complement of certain adpositions; used in a construction called the 'ablative absolute';
- vocative: the form of address;
- locative: for some place names and certain lexemes such as {RŪS} 'countryside' and {DOMUS} 'home', this form (distinct from the ablative form) encodes 'locative' meanings (i.e., 'in', 'at').[3]

[2] Woodcock (1959) and Pinkster (2015: chapter 12) offer more extensive discussion on the most common functions of the morphological case-forms in Latin.

The Latin case-forms as they are presented here actually appear to change very little through time when one considers the majority of the extant data. However, the conservative nature of the case system (and the language more generally) is a reflexion not of a lack of development but of the artificiality of the written medium. Despite abundant attestations of the language through time, therefore, tracing the decline of the case system accurately is very difficult, as the written data do not necessarily reflect changes that may have affected the spoken language.[4] The modern Romance languages show that the morphological expression of case found in Latin has been lost almost entirely lost from nouns and adjectives. What remains uncertain are the precise ways in which and reason(s) why these inflexional case distinctions were effaced diachronically.

The rest of the discussion in this section focuses on presenting some of the explanations that have already been offered for the changes exhibited by the Latin case system. It will be demonstrated that, although many details are captured by appealing to a network of phonological, analogical, syntactic, and contact-induced developments, the exact details of and reasons for the loss of the Latin case system remain unclear.

2.1.2 THE DEVELOPMENT OF THE LATIN CASE SYSTEM - PHONOLOGY

A major factor that is usually discussed in reference to the loss of the Latin morphological case distinctions in nominals is phonetic erosion.[5] Certainly, on consideration not only of data provided by the Romance languages, but also from inscriptions and other writings, there is

[3] The vocative does not encode 'syntactic' functions and is therefore not considered a 'case' here; it is not included in the discussion that follows. The locative will also be omitted because it is a fossilized form and not productive in the language.

[4] The written language also frequently fails to represent the stylistic and regional differences that were likely to have been found at all stages during which Latin was spoken. This is discussed by, for instance, Adams (2003, 2007, and 2013).

[5] Most authors first mention 'phonological changes' as one of the key elements in the development of the Latin case system, e.g., Herman (2000:27-48; 52f.), Clackson and Horrocks (2007:277).

plenty of evidence to support the hypothesis that phonological developments have played a part in the loss of the distinct marking of certain case values.

One change which is posited to have occurred at a relatively early stage in Latin is the loss of final *-m*.[6] The evidence for this change is extensive, and comes from various different sources, including: comments made by contemporary grammarians;[7] metrical scansion; inscriptional evidence; and consideration of inherited forms in the modern Romance varieties.[8] The loss of final *-m* affected all non-neuter accusative singular forms, certain nominative/accusative singular neuter forms, and genitive plural forms. In the first declension, as a result of this change, the nominative and accusative in the singular were no longer distinguished morphologically:

	Classical Latin	**Sound Change**	**Result**
Nominative singular	PUELLA	--	*puella*[9]
Accusative singular	PUELLAM	Loss of final *-m*	*puella*

Table 2.4: Loss of *-m* in Latin first declension singular nominals

The loss of final *-m* would also have opacified the formal differences between the accusative singular and ablative singular forms of non-neuter third declension nominals, e.g., accusative singular RĒGEM > **rēge*,[10] ablative singular RĒGE.

Furthermore, following the loss of length distinctions, first in unaccented and thereafter in accented syllables,[11] the ablative singular form of first declension nominals 'fell together' phonologically with the nominative/accusative singular. As a result of this change and the

[6] Via an intermediate stage of 'muffled' articulation which can be "traced well back into the Republic". Adams (2013:129).
[7] Adams (2013:128-132).
[8] There are a very few, monosyllabic words in some Romance languages which retain Latin *-m*, e.g., Latin (accusative singular) REM 'thing, matter' > French *rien* 'nothing'. Väänänen (1981:67).
[9] The forms in tables 2.4 and 2.5 omit from consideration phonological changes affecting the lexical stem.
[10] Here and throughout a single asterisk * is used to indicate that the form is etymological but not attested; a double asterisk ** indicates that the form is not etymological; the symbol † is used for nonce forms.
[11] Clackson and Horrocks (2007:272f.), Adams (2013:43-51).

loss of final *-m*, first declension Latin nominal would have had just two distinct forms in the singular paradigm:

Singular	Classical Latin	Sound Change	Result
Nominative	MENSA	--	**mensa*
Accusative	MENSAM	Loss of final *-m*	**mensa*
Genitive	MENSAE	*ae* > [e][12]	**mense*
Dative	MENSAE	*ae* > [e]	**mense*
Ablative	MENSĀ	Shortening of a long vowel in an open syllable	**mensa*

Table 2.5: Phonological developments affecting first declension singular nominals

Other developments affecting the vowel system led to further instances of syncretism in other declensions. For instance, original long *o* and short *u* merged as close /o̜/, as seen in old French and old Occitan (although not Romanian).[13] In the second declension, this led to the syncretism of the nominative singular and the accusative plural (SERUUS and SERUŌS > */*sɛrwo̜s/*).

Despite frequent instances of phonetic erosion which resulted in the syncretism of certain inflected forms, the near-total loss of the inflexional case system is not the result of phonological developments alone. For example, although the genitive plural forms *-ārum*, *-ōrum*, *-um*, *-ium*, and *-uum* would all have lost their final *-m*; the long *-ā* in the first declension genitive would no longer have been distinguished from short *-a*; the long *-ō* in the second declension exponent would have become /o̜/, none of the changes would have led to the genitive plural forms *-ārum* and *-ōrum* falling together phonologically with any other case-form in the first and second declensions respectively. There are many other instances besides, where regular phonological developments would certainly have altered the Classical

[12] Herman (2000:31), Adams (2013:71-81).
[13] Väänänen (1966:27) and (1981:36f.), Herman (2000:31-33).

forms of the Latin case system, but would not have resulted in a single, invariant form in either the singular or the plural.[14]

2.1.3 THE DEVELOPMENT OF THE LATIN CASE SYSTEM - ANALOGY

Certain changes affecting the case system are analysed as analogical in nature. For instance, in inscriptions from Africa and Gaul, we find first declension nominal forms ending in *-ās* which syntactically encode nominative plural functions:[15]

hic quescunt duas matres duas filias

'here lie two mothers (and) two daughters'[16]

The change *ae* > *ās* is not the result of regular phonological change. It is suggested instead that there was an extension of the pattern [nom. pl.] = [acc. pl.] found in third declension nominals, which may have been facilitated by the structural similarity of the third declension nominative/accusative plural exponent *-ēs* in non-neuter lexemes (where both *-ās* and *-ēs* consist of a long vowel followed by a final sibilant).[17] This development of non-etymological syncretism accounts for the inscriptional data found particularly in Africa and Gaul,[18] and may help explain why the old Gallo-Romance feminine nominals which have developed from first declension Latin nominals do not distinguish case morphologically in the plural.

Second declension nominals were characterized by an accusative plural form ending in *-ōs* - that is, a long vowel followed by a final -s, which thus exhibits phonological similarities to

[14] Unlike case, morphological number (i.e., 'singular' vs. 'plural') distinctions are robustly maintained throughout Romance. Maiden (2016c).
[15] Adams (2013:251f.).
[16] This inscription is discussed by Väänänen (1966:84), (1981:108), Gaeng (1977:46-51), and Adams (2007:674f.), (2013:251) amongst others. There may also be earlier examples in Atellan farce. Adams (2013:251).
[17] The pattern [nom. pl.] = [acc. pl.] is also found in the fourth and fifth declensions, similarly with the shape of long vowel followed by final -s, which may have afforded extra paradigmatic pressure on first declension nominative plurals.
[18] There is also evidence to suggest that the nominative plural ending in *-ās* occurred in Italo-Romance. Maiden (1996).

the accusative plural endings of the first (*-ās*) and non-neuter third (*-ēs*) declensions. However, there are only rare instances of second declension nominals with a nominative plural form in *-ōs*,[19] and it is not clear why these lexemes should continue to maintain case distinctions in old Gallo-Romance whilst inherited first declension nominals do not. Taking a network view of the data, however, and considering the effects of language contact in addition to analogical pressure, I will offer a potential reason for the retention of the nominative plural *-ās* and the lack of nominative plural *-ōs* in §2.1.5.

In addition to the development of first declension accusative plural forms in *-ās*, the loss of certain inflexional patterns may also be described in terms of analogy. In Latin, there are five major declensions, although the fourth and fifth included fewer lexemes than the first, second, and third, and were not productive in the classical language; these two more marginal classes have been almost entirely lost from the Romance languages. Possibly, the formal similarity of non-neuter nominative singular (in *-us*) and accusative singular (in *-um*) forms in the second and fourth declensions may have aided the incorporation of fourth declension lexemes into the second declension. For instance, where a language-user knows only the nominative singular and/or accusative singular forms of a lexeme, there is not sufficient information to determine whether the nominal belongs to the second or fourth declension (although any other form of the lexeme is diagnostic of the inflexion class).

More frequent lexical items, of which a greater number of inflected forms are more likely to be known, could possibly have maintained the inflexional patterns found in the fourth declension. Analogy of the form~function variation of the nominative singular and accusative singular forms of the second and fourth declension lexemes, therefore, does not necessarily

[19] Herman (2000:72f.), Adams (2013:252).

explain the loss of the fourth declension.[20] However, when considered in combination with a 'rise in configurationality' (discussed in §2.1.4), this analogical explanation may acquire extra force. That is: speakers began to prefer prepositional phrases over the bare genitive, dative, and ablative forms in the distinguishing of particular functions. As a result, the inflexional forms even of more frequent lexemes generally would have been encountered and used less and less often. It is possible that a language-user, over time, would have in fewer instances been familiar with an inflected form other than the nominative or the accusative. Thus, the formal similarity between the nominative and accusative in the singular may well have been, at least in part, the basis for the transfer of lexemes from the smaller inflexion class to the larger.

Even this altered explanation, however, does not account for the loss of distinct second and fourth declensions in the Romance languages, as it is only in the *singular* in that the nominative and accusative exhibit identical inflexional behaviour in Classical Latin:

	Second Declension		**Fourth Declension**	
	Plural		**Plural**	
Nominative	SERVĪ →	**servi*	GRADŪS →	**gradus*
Accusative	SERVŌS →	**servọs*	GRADŪS →	**gradus*

Table 2.6: Partial Latin second and fourth declension plural inflexional patterns

Whilst long /o:/ and short /u/ would have fallen together phonologically, long /u:/ developed into /u/ in Romance, meaning that plural forms across declension would have remained morphologically distinct. Thus, knowledge of either the nominative plural or the accusative plural form would have enabled a language-user to determine whether a nominal belonged to the second or fourth declension.

[20] There remain some fossilized relics of the latter declension are found in certain central and southern Italian dialects, such as *mano* 'hand' and *fico* 'fig', both invariant for number. Maiden (1996).

The transfer of fifth declension lexemes to the first declension poses further problems for a strictly analogical approach, since the morphological case/number exponents of these two inflexion classes are in no instances the same. It may instead be suggested that the gender of the nouns belonging to these inflexion classes (since the majority of nominals in both were feminine, with very few masculine and no neuter members) may have facilitated the transfer of lexemes from the unproductive fifth declension to the larger first declension. We also find feminine fourth declension nouns allocated to the first declension (e.g., SOCRU(M) 'mother-in-law' > Spanish *suegra*, Italian *suocera*, Romanian *soacră*),[21] potentially lending some support to the hypothesis that gender can aid in the transfer of lexemes across inflexion classes.

Furthermore, although there are no instances in which the same suffixal forms of first and fifth declension lexemes differentiate precisely the same case/number values, there is a similarity in the patterned distribution of forms, where in both of these inflexion classes the genitive singular form was identical to that of the dative singular. This sort of patterning, in addition to considerations of gender, may have supported the transfer of nominals from the smaller and unproductive fifth declension into the larger and more productive first.

The loss of the fourth and fifth declensions would not have led to the loss of particular case values, as the five cases counted in §2.1.1 would still have been differentiated morphologically in at least one of the three major declensions (even following the analogical and phonological developments outlined here and in §2.1.2). However, the loss of the inflexional patterns exhibited by nominals belonging to the smaller inflexional classes meant that the number of distinct desinences which distinguished, for instance, 'ablative singular' or 'genitive plural' functions, was reduced.

[21] Maiden (2011a:163).

Although analogy played a part in certain changes in the inflexional behaviour of certain Latin lexemes, it did not in and of itself lead to the loss of particular case values. When the analogical developments discussed here are considered alongside other developments (such as phonetic erosion and the 'rise of configurationality'), though, we may better account for certain changes found to have occurred in the Latin case system through time.

2.1.4 THE DEVELOPMENT OF THE ITALIC CASE SYSTEMS - SYNTAX

Alongside phonological and analogical factors, certain syntactic developments which occurred in the history of Latin are often cited as complementary causes for the loss of the case system. These include: the increase in analytic case-marking, the rise of configurationality, and the gradual replacement of 'dependent-' with 'head-'marking.

The preference for synthetic or analytic means of expressing particular functions is often recognized as an important difference between Latin and the modern Romance languages. In the case system, the difference is striking: in Latin, many case functions were encoded by morphological alternations which affected the ending of the nominal; the expression of case (and number) was inseparable from the lexical stem. Omitting Romanian from consideration, no modern Romance variety expresses case this way in its nominals. Instead, analytic means - particularly prepositions, but also clitics - are used.

Latin also employed analytic means of expressing particular functions, and it has been claimed that the loss of morphological case was facilitated by the existence of prepositions which encoded the same meanings as could be expressed morphologically. For instance, Herman observes that the preposition AD 'towards' and a nominal inflected in the accusative could express the indirect object, and that this construction was around even from an early stage of Latin. He cites examples such as the following from Plautus:

2.1.a QUAE AD PATREM UIS NUNTIARI

which to father.acc you.want to.send

2.1.b NUMQUID ALIUD UIS PATRI NUNTIARI?

Surely other you.want father.dat to.send[22]

However, it does not seem as though these two alternatives were actually functionally equivalent in Latin until the medieval period: instead, the prepositional phrase "conveys the idea of transportation towards someone, whereas the dative encodes the sense 'to communicate to'."[23] Thus, the prepositional phrase encodes a more spatial meaning rather than a strictly grammatical one.[24] Since these two are not semantically equivalent, it is difficult to see how the prepositional phrase could have simply 'taken over' from the synthetic means of expressing the indirect object.

Throughout the history of Latin, however, prepositions such as AD come to be used instead of case-forms in the expression of particular functions, even though the functions of the synthetic and analytic constructions were not absolutely equivalent in Latin. Some scholars, such as Penny, have claimed that the reason for this emergent preference for prepositional marking of case was because "the inadequate distinction existing in early Latin between one case and another led to the use of prepositions in a disambiguating function, a development which made the case-endings largely redundant and allowed further merger, as a result of phonological change, to go ahead unchecked."[25] Thus, where phonological developments

[22] Examples from Herman (2000:61).
[23] Pinkster (1990:202).
[24] In language more generally, it is supposed that different forms are associated with slightly different functions, whether these have to do with semantic, grammatical, or even diastratic or diatopic distinctions, for instance (as will be discussed in §4.1.2.) It is therefore probable that these two means of expressing the 'indirect object' were perceived as in some way functionally distinct by native speakers, even if it is not clear to us now the ways in which the two construction types were distributed.
[25] 2002:117. Similarly, Väänänen (1981:112) says in reference particularly to the accusative and ablative that "des tours prépositionnels se substituent par besoin de clarté aux formes casuelles pures et simples".

opacified the morphological distinctions between certain case values, speakers began to rely on prepositional means to distinguish particular functions. For example, the preposition IN could take a complement inflected in either the accusative or the ablative; the difference in meaning was one of 'goal' vs. 'stasis': IN AQUAM (accusative singular) translates as 'into the water', whilst IN AQUĀ (ablative singular) translates as 'in the water'. However, following the developments described in §2.1.2 (namely, the loss of word-final *-m* and the loss of phonemically distinctive lengths in unstressed vowels), the prepositional phrase **in aqua* could have meant either 'into the water' or 'in the water'. In order to distinguish these functions, phonologically distinct prepositions might be used.

There are two main counter-arguments to the claim that prepositions played a disambiguating role as the morphological distinctions between certain case-forms became blurred. First, context would have, in very many instances, provided sufficient information to correctly determine whether a construction such as **in aqua* encoded goal or stasis. Second, the idea that prepositions played a 'disambiguating' role where morphological case forms encoded more than one function is made problematic by consideration of the frequent polyvalency of many modern Romance prepositions. In some modern Romance varieties, distinct prepositions are used to differentiate functions that would have been encoded by a single inflexional form in Latin (i.e., the ablative). For instance, in Romanian:

2.2.a *lampa este pe podea*

'The lamp is on the floor' (locative)

2.2.b *vine din străinătate*

'(S)he comes from abroad' (ablative)

2.2.c *citeşte cu ochelarii*

'(S)he reads with glasses' (instrumental)[26]

This sort of form~function isomorphism is rare and, in most Romance languages (including Romanian), prepositions tend to encode more than one function. One striking example comes from modern French, where the preposition *de* can be used to express a similar variety of functions as the Classical Latin genitive case:

Construction	**Latin**	**French**	**Gloss**
N. complement	REX REGUM	*le roi des rois*	the king of kings
A. complement	AUIDUS GLORIAE	*avide de gloire*	eager for glory
V. complement	MEMINI UIUORUM	*je me souviens des vivants*	I remember the living
Partitive	MULTI CIUIUM	*beaucoup des citoyens*	many of the citizens
Quality	UIR MAGNAE ELOQUENTIAE	*un homme d'une grande éloquence*	a man of great eloquence
Possessive	DOMUS REGIS	*le palais du roi*	the king's palace

Table 2.7: Latin genitive and French *de* compared[27]

Since polyvalent prepositions are very common across Romance (as well as in Latin), it cannot be argued that prepositions always played a disambiguating function as the distinct Latin case-forms were lost. Such simplistic 'replacement' of inflexional with syntactic case-marking does not lend explanatory force to the changes observed.

In this study, the development towards a preference for analytic means of expressing case which occurred through time is considered to fall within the scope of the rise of configurationality.[28] Configurationality has to do with the fixed order of constituents which permits a language-user to correctly express and interpret the syntactic functions these constituents encode. Whilst word order in the Romance languages is governed by "syntactic

[26] Examples adapted from Nedelcu (2013d:453).
[27] Adapted from Vincent and Börjars (2010).
[28] After Ledgeway (2012).

factors [and] by considerations of discourse and information structure",[29] in Latin, word order was governed by pragmatic considerations.[30] This increased reliance on word order to distinguish certain syntactic functions marks an increased reliance on configurationality through time. Thus, in both a configurational and a synthetic~analytic approach, the differences between Latin and the (modern) Romance languages have to do with the changing ways in which the relationships between linguistic constituents are encoded, that is, whether by non-configurational or configurational syntax.[31]

The development of head- vs. dependent-marking is similarly considered to fall within the scope of the rise of configurationality. The move from dependent- to head-marking is one that can clearly be seen in the shift from morphological to syntactic marking of case. In a Latin sentence such as MISIT EPISTULAM 's/he sent the/a letter', for instance, the dependent noun EPISTULAM is morphologically marked for the relationship it bears to the head MISIT. In a head-marking language such as (modern) French, on the other hand, we find examples such as 'il/elle a envoyé *une lettre*', where word order is the primary means by which the subject (underlined) is differentiated from the direct object (italics); thus, these functions are distinguished by the position of the dependent noun in relation not only to the head (the verb) but also other constituents in the sentence. Whilst it cannot be argued that Latin is fully 'dependent-marking' or that the Romance languages are fully 'head-marking', Latin certainly had more dependent-marking constructions than Romance.[32]

Although the rise of configurationality and the loss of the Latin case system are no doubt intertwined developments, it is difficult (perhaps impossible) to argue that either: a) the rise of configurationality *caused* the loss of morphological case distinctions, or b) the loss of

[29] Vincent (2016:38).
[30] Cruschina (2016:596).
[31] Ledgeway (2012:287f.).
[32] This accords with the cross-linguistics tendency for languages to more consistently employ one type over another. Nichols (1986:56).

morphological case distinctions *caused* the rise of configurationality.[33] It is certainly possible that phonological developments led to a greater need for other means (word order, prepositions, clitics) of differentiating certain functions, and the rise of configurationality as an explanatory factor in the loss of the Latin case system is one that is best understood from a network perspective, in combination with the many other developments that have already been discussed. However, the exact ways in which these factors interacted are unknown (and likely to remain that way). Whilst an evolutionary perspective is essential, therefore, it does not provide all the answers to our questions given an unfortunate lack of data.

Both old Gallo-Romance and Romanian exhibit synthetic and analytic means of expressing particular functions which appear only at first to be equivalent. Further detail will not be offered in reference to these varieties about the syntactic means of expressing case, nor about the rise of configurationality, as similar observations as those offered here in reference to Latin pertain *mutatis mutandis* to these varieties.

2.1.5 THE DEVELOPMENT OF THE ITALIC CASE SYSTEMS - CONTACT CONSIDERATIONS

Even from the earliest days of its history - before the spread of the Empire - Latin was in contact with other languages, including its close relatives Umbrian and Oscan, more distantly-related members of the Indo-European family (e.g., Ancient Greek), and unrelated varieties (e.g., the language isolate Etruscan). As the Empire extended its reach, speakers of Latin started to come into contact with even more different linguistic varieties.[34] It is known that contact with other varieties can affect languages in a number of ways. In addition to the borrowing of lexical items, there is evidence to suggest that other levels of interference - including phonological, morphological, and syntactic, for instance - also occur in contact

[33] There are languages with cases and fixed word order (e.g., Icelandic) as well as languages with little to no case morphology that nevertheless exhibit relatively free word order (e.g., Bulgarian). Butt (2006:4f.).

[34] Adams (2007:245-284) offers discussion on the evidence for contact with Gaulish, Punic, Libyan, Berber, Aramaic, Hebrew, Germanic, Egyptian, Getic, Sarmatian, and Thracian, as well as the three 'main' Hispanic languages Iberian, Celtiberian, and Lusitanian.

situations (although these tend to be more difficult to detect than loanwords). This section will investigate whether the loss of the Latin case system can be attributed to contact with other languages.

According to Trudgill (2011), there are two main types of developments that can take place through time: 'complexification' and 'simplification'.[35] The first of these tends to occur in those linguistic communities which are either isolated from other languages, or which are in long-term, socially-stable contact situations. 'Simplification' on the other hand is often found in less socially-stable contact situations in which there is a great deal of short-term L2 learning. The loss of an inflexional category is considered by Trudgill to constitute an example of 'simplification', and there is evidence to suggest that Latin has been in contact situations in which simplification is expected to take place. For instance, although the Romans did not exert a despotic pressure on foreigners to learn Latin, tradespeople and those who wished to 'rise through the ranks' of the social strata recognized the importance of learning (although perhaps to different extents and for different purposes) the language of the invaders. In the earlier years of the Roman occupation in the provinces, therefore, there may have been more imperfect adult learning than native acquisition, which may have led to the 'simplification' of the case system as a result of the L2 preference for analytic marking.[36] The rise of configurationality (§2.1.4) may have therefore been the result of the interplay of both internal *and* external pressures.

To discuss in detail the particulars of the sociolinguistic situations in which Latin was used throughout the Empire is made extremely difficult owing to the paucity of the data. However, a few details can be noted in reference to the development of Latin as it was spoken in Gaul

[35] I am uneasy about the terms 'simplification' and 'complexification', although Trudgill goes to great lengths to assure the reader that the changes he discusses are in fact instances of either simplification or complexification in language change. When these terms are used throughout, scare quotes will be used to indicate that they are in some ways problematic.
[36] Thomason and Kaufman (1988:55).

and in the Balkans, where old Gallo-Romance and Romanian respectively developed. Before I do so, it is important to note that, although we have knowledge of at least some of the languages with which old Gallo-Romance and Romanian might have been in contact, this knowledge is in many instances very limited. For example, very often we do not have extensive understanding of the grammars of these varieties, nor of the ways in which speakers of Latin perceived these languages (e.g., whether they were considered 'fashionable' or 'learnèd'), nor of who exactly learned them (whether tradespeople, the upper classes, the lower classes, people living in certain geographical areas, etc.). All these different factors and more can have an effect on the ways in which languages in contact develop; not having this sort of information means that any observations offered here are tentative at best.

When Latin was brought to Gaul, it came into contact with the Celtic language Gaulish, and the results of this contact are most clearly seen in the many loanwords borrowed from Gaulish into Latin, particularly those to do with transport and horsemanship.[37] There are also a small number of bilingual inscriptions which permit us some insight into some of the grammatical structures of Gaulish. Most notably for our purposes, Gaulish was characterized by an inflexional case system (although it is not known for certain how many values were distinguished morphologically; nominative, accusative, genitive, and dative values are usually listed in traditional grammars).[38] Our understanding of Gaulish is, however, much more limited than that of Latin, which means that any observations regarding the changes which may have been induced as a result of contact with this Celtic language must, by necessity, be hesitant. However, one example of a contact-induced change affecting the Latin case system may be found from the first century AD in

> "examples of [nominative plural] *-as* in the graffiti of La Graufesenque, in both Gaulish and Latin texts, in lists of objects in which, at this site, the

[37] Palmer (1954:53).
[38] A vocative is also commonly listed. Dottin (1918:117).

> nominative is the usual case. The ending of the nominative plural in Gaulish was *-as*, and at the pottery that morpheme was transferred by the Gaulish potters into their Latin."[39]

Although Adams is at pains to stress that this is not necessarily the only source of the nominative plural *-as* (generally analysed as the origin of the unexpected *-s* in old French and old Occitan feminine nominative plural lexemes from the Latin first declension), it may have aided the analogical extension of the Latin accusative plural form *-ās* to encode nominative plural functions (as discussed in §2.1.3). It may also explain why this pattern of syncretism persisted into old French and old Occitan, whilst the nominative plural and accusative plural forms of inherited second declension nominals remained distinct, since there was no nominative plural form in *-ōs*.

Gaul was lost in the fifth century to various Germanic-speaking tribes,

> "when the south was occupied by the Visigoths, the eastern central area by the Burgundians and the north by the Franks. The Franks were the most powerful of these Germanic tribes and their king, Clovis (c. 465-511), succeeded in extending his rule over most of Gaul."[40]

These Germanic languages were characterized by morphological case. Intriguingly, however, unlike in the old Gallo-Romance/Gaulish contact situation, speakers of old French did not take up the language of their conquerors; instead, the Franks actually adopted old French and abandoned their native tongue.[41] This may present an example of a contact environment which led to 'complexification' - or at least to the retention of original 'complex' morphology of the case system - where the two languages were in close and long-term contact, and thus old French was learned natively by the children of the Frankish invaders. Such a suggestion must, however, remain speculative.

[39] Adams (2007:675).
[40] Price (1971:4).
[41] Price (1971:4).

Considering the effects of the contact situation in which Romanian has been spoken and learned through time is very difficult given the lack of information about the nature, intensity, and time-depth of Romanian's contact with other linguistic varieties. There is lexical, typonymic, hydronymic, and historical evidence to suggest that Romanian has been in contact with many different languages throughout history, but direct evidence of these - from inscriptional evidence or other writings, for instance - is rare. Even more problematically, there is evidence of languages in the Balkans which are *not* directly attested by extant written data. For instance, Thraco-Dacian, an Indo-European language, is referred to by various ancient writers; occasionally, proper names (including those of people and places) are found on coins or in Greek and Latin inscriptions;[42] and it is likely that Latin was spoken in and around the areas in which Thraco-Dacian was also used. It can therefore be concluded with reasonable certainty that they were in contact, but any more concrete conclusions about the effects that this distantly-related variety may have had on Romanian - and particularly on its case system - are impossible to offer.

Although many of the languages with which Romanian was in contact are directly attested (in contrast to Thraco-Dacian), it is not always clear exactly what type of influence these languages may have had on the development of Romanian. Pană Dindelegan asserts that "[t]he Latin structure of Romanian has not been modified" by contact with neighbouring varieties, and that "all influences [are] limited to lexis and word formation."[43] However, this may be a reflexion of the limited nature of the resources from which we may draw conclusions, rather than a true indication of the effects of language contact.

In terms of the case system, there are a couple of things to note. In the first place, Romanian has long been in contact with many other varieties which maintain robust case

[42] Pană Dindelegan (2013b:2f.).
[43] 2013b:3.

systems, including Hungarian, Albanian, Greek, and various Slavonic varieties. This long-term and frequently intense contact with other varieties which maintain morphological case distinctions may in some way have aided the maintenance (or perhaps the innovation - §2.3.2-3) of case distinctions in Romanian. Although it is possible for morphological markers to get borrowed in contact situations, it is very rare, and only really found in those instances where there is "very extensive bilingualism among borrowing-language speakers, [and] social factors strongly favoring borrowing".[44] Since the vocative marker *-o* has been borrowed from Slavonic,[45] it is not unreasonable to suggest that the linguistic variety has affected Romanian in many other significant ways, and this might include changes to the case system.[46] Because of the paucity of evidence, though, it is unlikely that any firm conclusions will ever be offered about these effects.

In order to offer more specific insights about certain developments that occurred in Latin, old Gallo-Romance, and Romanian, it would be essential to know in detail which languages these varieties were in contact with. Such information is, unfortunately, limited. Any appeal to contact as a causal factor in the loss of the Italic case system in nominals is, therefore, at best speculative. Here, too, is found a weakness of the application of the evolutionary perspective to language change, since an accurate account of diachronic developments is reliant on a great deal of data not only on phonological, analogical, functional, and syntactic developments, but also the pressures exerted by social considerations, which include, but are not limited to, contact situations and the complex attitudes towards different speech communities. Since recent work (by, e.g., Trudgill 2011 and Gardani 2018) has demonstrated that contact can have far-reaching effects on all the languages in a given contact situation, it is essential, to include this consideration in our conclusions, though with due care.

[44] Thomason (2001:70f.), Gardani (2018).
[45] Nedelcu (2013b:273).
[46] N.B., I do not here consider the Romanian vocative a case.

2.2.0 INTRODUCTION TO OLD GALLO-ROMANCE

At an earlier stage of the Gallo-Romance varieties a two-value case 'system' is found in nouns and adjectives - a distinction which has since been lost. It occurs in the earliest writings (from the ninth century for early northern Gallo-Romance) and is considered to have been retained in old French until roughly the end of the thirteenth century in the spoken register (perhaps a century longer in writing) and in old Occitan until the beginning of the thirteenth century.[47] In the following sections, I will describe the old French and old Occitan case systems as they are presented in synchronic grammars of the languages, though it will be stressed that these are 'idealizations' of the data and that the case system as it is found in attested literature is in many ways much more complicated than accounted for by prescriptive grammarians. I will also offer an account of the diachronic developments of these case systems.

2.2.1 THE OLD GALLO-ROMANCE CASE 'SYSTEM'

The nominative forms in old French and old Occitan descend phonologically from the Latin nominative form; the oblique (mostly) from the Latin accusative form.[48] In addition, two gender values (masculine and feminine) survive from Latin, and 'singular' and 'plural' values are distinguished in most lexemes. In both old French and old Occitan, case is more consistently distinguished in masculine nominals in both the singular and the plural than in feminine nominals.

The functions of the nominative and oblique cases are usually considered to be very similar in old French and old Occitan:

[47] Kihm (2017:41).
[48] "[T]hough it would perhaps be more accurate to consider [the old French oblique], at least in the singular, as representing a fusion of the accusative and the ablative." Price (1971:94); also Salvi (2011:319).

	Case	Functions
Old French	**Nominative**	The form used to encode the subject and the vocative.
	Oblique	Used to encode object functions; the form of the complement of prepositions; can express possession; occasionally the form of the indirect object.[49]
Old Occitan	**Nominative**	The form used to encode the subject.
	Oblique	Used to encode direct objects; the objects of prepositions; possessive and indirect object functions; in adverbial phrases without a preposition; in an 'absolute' construction.[50]

Table 2.8: Functions of the old Gallo-Romance nominative and oblique cases

These are the traditional functions listed for the nominative and oblique forms in old French and old Occitan, but a number of scholars have noted that attested uses of inflected nominals present a picture not quite as clear-cut as that offered by grammarians. For instance, in old French, "some syntactic structures [in the *Chanson de Roland*], particularly those in apposition or lacking an explicit predicate, tend to favour the oblique case-form over the subject form".[51] Additionally, Woledge et al. observe that metrical considerations may in some instances take precedence over syntactic ones in the selection of particular case-forms,[52] and even geography plays a part in determining how nominative and oblique forms are distributed in a text.[53] In fact, it has famously been noted that "si l'on met à part les plus anciens textes, ceux du IX^{e} et du X^{e} siècle [...], les règles de la déclinaison n'apparaissent dans toute leur pureté que dans les grammaires modernes de l'ancien français".[54] Where different morphological forms were used, Smith claims that "it was as a grammatically unmotivated marker of archaism".[55]

[49] Nyrop (1924:127-141), Foulet (1970), and Moignet (1976).
[50] Paden (1998:281-287). Both the nominative and oblique forms could be used for address. Jensen (1976:127).
[51] Sornicola (2011:26), referring to work conducted by Woledge et al. (1967).
[52] 1967:166f..
[53] van Reenen and Schøsler (1988:508-512, 523-526).
[54] Bédier (1927:248).
[55] 2016:303.

To account for the old French data, Schøsler (1984) argues that the old French case system did not function in quite the same way as that of Latin. That is, the nominative and accusative forms were not always used to consistently 'encode' specific syntactic functions. Instead,

> "[i]l faut donc constater que, seuls parmi les diverses fonctions nominales, le prime et le second actant ne semblent signalés qu'à l'aide de la déclinaison bicasuelle: *distinguer formellement le prime et le second actant, cela doit être la fonction essentielle de la déclinaison*."[56]

In many instances, particularly in sentences with one animate and one inanimate referent, the context would have been sufficient to determine which was the 'prime actant' - that is, the subject of the verb - and which the 'second actant' - the direct object. However, where two masculine nouns, both with animate referents, are found, inflexional variation could serve to differentiate syntactic functions where context and word order did not.

In reference to the old Occitan case system, it is similarly noted that "[t]owards the beginning of the 13th century [...], the two-case system is not strictly adhered to in all quarters"; in other dialects and areas, the 'vacillation' of case-forms is found as early as the eleventh century.[57] In many texts, therefore, we find what might be considered 'errors' in the distribution of certain forms according to the prescriptive ideal presented in grammars of old Occitan, which may indicate that these distinct forms were not used in the same way as the Latin case-forms - that is, to encode certain syntactic functions - but instead perhaps for stylistic flair (cf. Smith's 2016 idea of 'markers of archaism') or to avoid syntactic ambiguity (cf. Schøsler's 1984 analysis of old French).

The distinct morphological forms of old French and old Occitan nominals are not distributed in the same ways across the written data, contrary to the impression created by grammars on the languages. It is possible that different old French and old Occitan authors

[56] Schøsler (1984:33), emphasis original.
[57] Jensen (1976:123).

used the case-marking of these languages in distinct ways, which is part of the reason that scholars cannot always 'make sense' of the textual distribution of the distinct nominal forms. Thus, it appears as though the formal variation exhibited by nominals in old Gallo-Romance consists of the remnants of the Latin system, rather than a system in itself.[58] Speakers were clearly aware of the presence of distinct forms for many lexemes, but do not appear to have shared a clear consensus on how these forms were to be distributed.[59] This may present an example of 'weakness' in the learnability of the case system, and may have contributed to its loss (discussed in §6.2).

In a similar way to the functions distinguished by particular forms, the forms themselves differ across texts, whilst in prescriptive grammars, inflexion classes are much more clearly defined. In both old French and old Occitan, three declensions for inflecting masculine nominal are listed:[60]

	Old French					
	{MURS} 'wall'		**{PERE} 'father'**		**{LERRE} 'thief'**	
	Singular	**Plural**	**Singular**	**Plural**	**Singular**	**Plural**
Nominative	*murs*	*mur*	*pere(s)*	*pere*	*lerre*	*larron*
Oblique	*mur*	*murs*	*pere*	*peres*	*larron*	*larrons*
	Old Occitan					
	{MURS} 'wall'		**{FRAIRE} 'brother'**		**{TROBAIRE} 'troubadour'**[61]	
	Singular	**Plural**	**Singular**	**Plural**	**Singular**	**Plural**
Nominative	*murs*	*mur*	*fraire(s)*	*fraire*	*trobaire*	*trobadór*
Oblique	*mur*	*murs*	*fraire*	*fraires*	*trobadór*	*trobadórs*

Table 2.9: Old French and old Occitan masculine nominal paradigms

The *murs*-type inflexion class is the most common and productive in old French and old Occitan masculines.[62] In this declension, case is distinguished in both numbers. These

[58] Sornicola (2011:27).
[59] Stanovaïa (1993).
[60] Lexical stems ending in *-s* or *-z* are invariant for number and case, e.g., old French feminine *pais* 'peace' and masculine *nes* 'nose'; old Occitan feminine *patz* 'peace' and masculine *nas* 'nose' (although later in the history of Occitan "a plural [...] was made for such words by adding -es". Grandgent 1909:89).
[61] Skårup (1997:9f.) discusses the use of these accents in old Occitan.

nominals are mainly inherited from the second and fourth Latin declensions. In old French *murs*-type nominals with stems ending in a velar, a labial, a liquid *l*, or a dental, the relationship between the stem and the inflexional exponent is opacified by the final *-s* of the nominative singular and oblique plural (as in table 2.10). Furthermore, some forms are characterized by an alternation *-z/-Ø* - that is, by the alternation of a final affricate rather than a final sibilant.[63] Nevertheless, the distribution of these forms remains the same as that of other *murs*-type nominals and will be described as 'chiastic', because of the X-like pattern exhibited by the two different wordforms across the four different paradigm cells:

	{ARS} 'arc'		**{CHEVAUS} 'horse'**		**{JORZ} 'day'**	
	Singular	**Plural**	**Singular**		**Singular**	**Plural**
Nominative	*ar-s*	*arc*	*chevau-s*	*cheval*	*jor-z*	*jor(n)*
Oblique	*arc*	*ar-s*	*cheval*	*chevau-s*	*jor(n)*	*jor-z*

Table 2.10: Old French {ARC}, {CHEVAUS}, and {JORZ} 'day' fully declined

Grammars also list three inflexion classes for inflecting feminine nominals in old French and old Occitan:

	Old French					
	{FILLE} 'daughter'		**{FLOR} 'flower'**		**{SEUR} 'sister'**	
	Singular	**Plural**	**Singular**	**Plural**	**Singular**	**Plural**
Nominative	*fille*	*filles*	*flor(s)*[64]	*flors*	*seur*	*serors*
Oblique	*fille*	*filles*[65]	*flor*	*flors*	*seror*	*serors*
	Old Occitan					
	{MAIRE} 'mother'		**{NAU} 'ship'**		**{SÓR} 'sister'**	
	Singular	**Plural**	**Singular**	**Plural**	**Singular**	**Plural**
Nominative	*maire*	*maires*	*nau(s)*	*naus*	*sór*	*seròrs*
Oblique	*maire*	*maires*	*nau*	*naus*	*seròr*	*seròrs*

Table 2.11: Old French and old Occitan feminine nominal paradigms

[62] Zink (1989:9); Paden (1998:203).

[63] In old French, the affricate *-z* is found instead of *-s* when the Latin form ended in *-t*, *-d*, *-st*, *-rn*, *-nn*, *-ļ*, or *-ŋ*. Pope (1952:310f.). For instance: PARTES 'parts' > *parz* (/parts/); HOSTES 'enemies' > ***ǫst-s* > *oz* (/ǫts/); DIURNOS 'daily' > ***džorn-s* > *jortz* (/*dzorts/*); **trapalios* 'work' > ***travaļ-s* > *travailz* (/*travaļts/*). In old Occitan, the affricate *-z* occurs where *-s* followed a lexical stem ending in a palatal or an undropped *-n-*, e.g., ANNŌS 'years' (accusative plural) > *anz*.

[64] "Il faut toutefois noter l'instabilité du *-s* désinentiel au singulier de ce type qui tend très tôt à s'aligner sur *fille*". Zink (1989:19).

[65] All feminine nouns and adjectives ending in /ę/ belong to this declension. Pope (1952:312).

Of the feminine nominals, only those belonging to the imparisyllabic declension faithfully maintain a distinction between nominative and oblique (although only in the singular). Number tends to be more faithfully differentiated morphologically by means of a final *-s*.

The declension classes given in tables 2.9 and 2.11 include both nouns and adjectives, the latter of which may be further split into two categories: those which distinguish gender and those which do not.

	Old French {BON} 'good'			
	Masculine		**Feminine**	
	Singular	**Plural**	**Singular**	**Plural**
Nominative	*bons*	*bon*	*bone*	*bones*
Oblique	*bon*	*bons*	*bone*	*bones*
	Old Occitan {BELS} 'beautiful'			
	Masculine		**Feminine**	
	Singular	**Plural**	**Singular**	**Plural**
Nominative	*bels*	*bel*	*bela*	*belas*
Oblique	*bel*	*bels*	*bela*	*belas*

Table 2.12: Old French and old Occitan gender-distinguishing adjectives

In the textual data, there are many instances in which a lexeme exhibits unetymological inflexional variation. This will be discussed in more detail in the sections that follow. Although it will be necessary occasionally to refer to certain nominal types (declensions or inflexion classes), it is to be remembered that the lexemes do not always inflect in the 'expected' ways in the textual data.

The loss of case distinctions in old Gallo-Romance has been attributed to a number of factors. In the first place, the apparent 'complexity' of the so-called case 'system' may have "ensured hesitation and confusion [...] and led to [its] demise."[66] Certainly, since the old Gallo-Romance declensions do not constitute a coherent system, but instead the relics of "an older array of forms which have been extended and remodelled in different ways according to

[66] Smith (2016:303).

place and textual traditions",[67] vacillation in the distribution of these forms was to be expected. This fluctuation, rather than terminating in a decided functional distribution of 'subject' and 'direct object' marking, for instance, ends in the modern Gallo-Romance varieties in formal invariance. Further to the consideration of 'complexity', scholars have posited phonological and analogical changes which may help to explain the total loss of the old Gallo-Romance case system, each of which will be explored in the sections following.

2.2.2 THE DEVELOPMENT OF THE OLD GALLO-ROMANCE CASE-FORMS - PHONOLOGY

As previously mentioned, in old French and old Occitan, the nominative forms have developed from the old Latin nominative; the oblique forms have developed largely from the Latin accusative. In the largest masculine inflexion class, these developments are frequently phonologically regular:

	Singular	**Plural**
Nominative	MŪRUS > *murs*	MŪRĪ > *mur*
Oblique	MŪRU(M) > *mur*	MŪRŌS > *murs*

Table 2.13: Old Gallo-Romance {MURS} from Latin {MŪRUS} 'wall'

The final *-m* found in the oblique singular form was lost early;[68] the loss of the unstressed vowels in each of the forms in table 2.13 leaves the final *-s* in the nominative singular and oblique plural forms and results in a pattern of syncretism which was not found in Latin.[69] Most of the lexemes in this inflexion class in old French and old Occitan descend historically, via regular phonological developments, from second declension Latin nominals.

Not all old Gallo-Romance forms have developed in this way, however. For instance, regular phonological developments cannot account for the alternating final sibilant found in

[67] Sornicola (2011:27).
[68] This phonological development is discussed in §2.1.2.
[69] Unstressed *a* became /ə/ in old French in most circumstances; all other unstressed vowels disappeared, "unless required to facilitate the articulation of preceding or following groups of consonants", before the ninth century. Pope (1952:112f.). Exceptions are given by Pope (1952:113f.) and Price (1971:58f.).

nominative singular forms and the lack of final sibilant in nominative plural forms of *pere(s)*-type (old French) and *paire(s)*-type (old Occitan) nominals. The nominative singular form of {PATER} 'father' in Latin ends in *-er* and not *-s*; moreover, the lack of *-s* in the oblique plural form is unexpected, since the Latin form PATRĒS would have encoded both nominative plural and accusative plural functions, yet only the oblique plural form in the old Gallo-Romance varieties consistently kept this final *-s*. These developments will be discussed further in §2.2.3.

The *fille/maire*-type inflexion classes continues mainly Latin first declension nominals as well as certain (feminine) forms from the Latin third declension (old Occitan *maire*, for instance, comes from the third declension Latin lexeme {MĀTER}, and could descend phonologically either from the nominative singular form MĀTER or from the accusative singular MĀTREM). In this large and productive inflexion class, case is not distinguished in either the singular or plural. Although the syncretism of *maire* is etymological (as this lexeme descends historically from a third declension Latin nominal in which nominative and accusative functions were not distinguished in the plural), the nominative plural and accusative plural are expected to be phonologically distinct in those lexemes from the Latin first declension. This syncretism is thought to be the result of the analogical extension of the accusative plural marker in (as already discussed in §2.1.3).

Nominals belonging to the old French and old Occitan 'imparisyllabic' declensions are generally inherited from Latin third declension nominals which exhibit a different number of syllables in the nominative singular (or nominative/accusative in neuter lexemes) when compared to the forms in the rest of the paradigm. In old French, lexemes such as {LERRE} are characterized by two different stem shapes:

	Singular	Plural
Nominative	LATRŌ > *lerre*	**latrōnī* > *larron*
Oblique	LATRŌNEM > *larron*	LATRŌNĒS > *larrons*

Table 2.14 Old French {LERRE} 'thief' from Latin {LATRŌ} 'mercenary soldier'

The case distinction found in the plural is not etymological, since the nominative and accusative functions in the plural of the lexeme {LATRŌ} would have been encoded by the same form in Classical Latin. There are two ways of accounting for this phonologically unexpected development. In table 2.14, a non-attested Latin form **latrōnī* is reconstructed which gives the old French form *larron* following the loss of phonemically distinct vowel lengths and the loss of short final *-i*; the exponent *-ī* is hypothesized to have been analogically extended from the nominative plural form of lexemes in the Latin second declension. Alternatively, the case distinction in the plural of masculine imparisyllabic forms may be analysed as the result of the analogical extension of the *-Ø/-s* alternation found in *murs*-type nominals. This will be discussed further in §2.2.3 and §6.2.

Subsequent developments affected the old Gallo-Romance case 'system', the result of which is the total lack of morphological case distinctions in modern French and modern Occitan nominals. As with the development of the case system from Latin, many consider phonetic erosion to have played a large part in this total loss. For instance, the regular development *s* ⟶ *Ø* / ____# C in old French affects those case-forms which are distinguished by the suffixation of a final sibilant. As a result of this, a form such as *pere* before a word beginning with a consonant would behave like an indeclinable nominal: the final *-s*, if there were one, would not have been heard and therefore would have played no disambiguating function about its role in the sentence. The forms and the functions they distinguished thus became opacified in the vast majority of nominals,[70] and it is estimated

[70] This also affected the morphological distinction of numbers in the majority of both masculine and feminine nominals.

"that at least one in three instances of inflectional *-s* was not realized phonetically in the spoken French of the thirteenth century."[71]

Despite the far-reaching effects of this phonological development, though, it does not sufficiently account for the total loss of case distinctions in old French. For instance, a stem alternation exhibited by old French nominals such as *sor ~ seror* and *lerre ~ larron* would not have been effaced by this development, so these forms should have maintained inflexional case distinctions in the singular into the modern language.[72]

Intriguingly, the old Occitan case 'system' has also been lost in the modern language, even though the language did not undergo the same loss of final *-s* as old French. The loss of case distinctions in this old Gallo-Romance variety therefore cannot be explained in the same way as has been suggested for old French. Jensen suggests instead that "[t]he gradual erosion and ultimate collapse of the declension were triggered by an intricate system of analogies",[73] which will be explored in more detail in the next section. Nevertheless, it remains clear that phonological developments alone cannot account for the total loss of the case system from old French and old Occitan, though they certainly had far-reaching consequences for the 'nominative' and 'oblique' distinction in old French.

2.2.3 THE DEVELOPMENT OF THE OLD GALLO-ROMANCE CASE-FORMS - ANALOGY

The previous section hinted at a number of analogical developments which may offer explanations for certain non-etymological forms found in old French and old Occitan. One of the most striking of these analogical developments is that of the (incomplete) extension of the *murs*-type chiastic distribution of the *-s/-Ø* suffix to *pere*-type nominals in old French and

[71] Smith (2016:303) citing Zink (1997:36).
[72] In the modern languages, there are occasional instances in which the nominative singular and oblique singular forms survive, although as distinct lexemes - for instance, (modern) French *on* 'one' (i.e., the indefinite subject pronoun) and *homme* 'man'. Smith (2005) and (2011).
[73] 1976:124.

paire-type nominals in old Occitan. As previously mentioned, the etymological form of the nominative singular of *pere*/*paire*-type nominals should not end in *-s*, whilst that of the nominative plural should. In the old Gallo-Romance data, however, we find frequent 'unexpected' forms. In particular, forms analysed as encoding 'nominative singular' functions ending in non-etymological *-s* co-occur alongside etymological forms *paire* and *pere* in the same syntactic environment. For instance, in old Occitan we find the following (different) nominative singular forms for the same lexeme:

altre(Ø) emfes (Appel 1930:13), but

e l'altres, cel de Flandres (Appel 1930:33)

The etymological nominative singular form from Latin ALTER 'other' should not exhibit final *-s* in the nominative singular, but in these examples we find instances of both etymological *altre* and non-etymological *altres* both apparently encoding subject functions. There are even occasional old Occitan examples of the vacillation of final *-s* in the nominative singular in the same passage:

vostre paire(Ø) fo mos compains (Brunel 1943:157)

These forms are not the result of regular phonological changes. Instead, the *-s* found in the nominative singular form and the lack of final sibilant in the nominative plural form might be analysed as the result of an analogical development, whereby the inflexional pattern exhibited by *murs*-type nominals has been extended to *pere/paire*-type nominals:[74]

[74] Pope (1952:313), Price (1971:95).

oblique singular : nominative singular

mur : murs

paire : ?

? = paires

Figure 2.1: Analogical development of non-etymological old Occitan nominative singular *paires*

The old Occitan nominative singular form *paires* (cf. etymological *paire*) has been analogically created on the implicational relationship existing between nominative singular and oblique singular forms such as *murs* ~ *mur*. Where nominative singular forms in non-etymological *-s* start to generalize, they might begin to be considered 'correct' by language-users, and therefore spread, perhaps to the detriment of the etymological forms. If this reflects what was happening in old French and old Occitan, however, the process is incomplete, and we find nominative singular forms such as *peres* and *pere*, *paires* and *paire* used alongside one another throughout the attested history of the old Gallo-Romance case system.

The analogical development of the nominative plural without final *-s* (where final *-s* is expected) appears to be more 'complete' than that of the nominative singular,[75] and has two possible origins. First, in the lexemes {PERE} and {PAIRE} it could be the result of the analogical extension of the Latin second declension nominative plural exponent *-ī* to third declension masculine nominals, which would have disappeared as a result of regular sound change and left a form in old French and old Occitan which did not end in a final sibilant. (This explanation also evidently accounts for the lack of *-s* in the nominative plural of lexemes etymologically inherited from the second declension.) The second possible origin of

[75] This 'completeness' of the nominative plural form without final *-s/-z* is more obvious in the prescriptive grammars; the actual textual data presents a much more complex picture of the situation.

the nominative plural in *-Ø* is that it is the result of the analogical extension of the *murs*-type pattern wherein the oblique plural form is characterized by a final *-s/-z* whilst the nominative plural form is not. This pattern is largely found exhibited by masculine lexemes, which may indicate that gender aided in its extension.

There are problems associated with the first of these explanations. In the first place, why would the second declension nominative plural exponent *-ī* generalize and re-establish a morphological distinction between 'nominative' and 'accusative' functions in third declension nominals, whilst the morphological distinction between these same functions in the first declension was analogically effaced? Moreover, if the syncretism of the nominative and accusative in the plural of third declension nominals was being disturbed, how could the pattern have been extended to first declension lexemes? In §6.4, it will be suggested that there is not sufficient evidence from Latin or Romance to support the reconstruction of third declension nominative plural forms ending in *-ī*. The second of the explanations offered above, it will be explained, is more likely to underlie the non-etymological nominative plural forms.

Finally, as was indicated at the end of §2.2.2, it is sometimes suggested that analogy was more important than phonetic erosion in the loss of the old Occitan case system.[76] In old Occitan masculine lexemes, the behaviour of final *-s* indicates that language-users closely associated it with nominative singular functions since, as discussed above, it is frequently found in nominative singular forms which etymologically should not end in a final sibilant. Jensen argues that the fluctuation of the nominative and accusative forms as the form of address "inevitably leads to a sporadic penetration of the nom[inative] into the domain of the acc[usative]".[77] This led to vacillation in the use of nominative and accusative case-forms,

[76] Jensen (1976:124).
[77] 1976:127.

particularly in certain constructions, such as ‘to be called’, ‘to become’, ‘to be seen’, as well as in comparative constructions. Thus, *-s* seemed to be less and less a marker of the nominative singular and, since it was in many instances a superfluous distinguisher of case (where context and word order were sufficient to differentiate ‘subject’ and ‘direct object’, for instance), it was dropped in favour of invariance. Alternatively, the more consistent syncretism of nominative and accusative in the plural of feminine nominals may have been analogically extended to masculine lexemes which maintained a morphological distinction between these case values in the plural.

There are a couple of problems with appealing to analogy as an explanatory reason for the total loss of the old Occitan case system. In the first place, why did the chiastic pattern of case-marking, as found in *murs*-type nominals, not generalize absolutely at the expense of lexical invariance, at least in masculine nouns? The analogical draw was clearly powerful enough to affect the plural forms of masculine nominals quite generally, and even the singular forms of certain *paire*-type nominals, so why is this type of case-marking not continued in the modern language? In the second place, it might be argued that the formal invariance within number found in the paradigms of feminine lexemes affected the masculines in that language-users analogically extended these patterns of invariance to *murs*-, *paire*-, and *trobaire*-type nominals. But why should the feminines have exerted such an influence on the masculine forms, and not vice versa?

Appealing to analogy may shed some light particularly on the introduction of non-etymological *-s* to certain forms in a lexeme’s paradigm. However, the analogical extension of inflexional patterns alone does not provide absolutely satisfactory reasons for the total loss of case distinctions in old French or old Occitan, since it is not always clear why one inflexional pattern should have generalized at the expense of another. Alongside phonological factors, however, analogical explanations afford us a fuller picture of the intertwined

developments which may have to some degree affected the forms and/or functions of the old Gallo-Romance nominative and oblique cases.

2.3.0 INTRODUCTION TO ROMANIAN

Romanian is the only modern Romance language that still maintains a case system in its nominals, although only in feminine singulars.[78] As in the previous sections on Latin and old Gallo-Romance, here I will first present a synchronic account of the (modern) Romanian case system (§2.3.1) and thereafter offer an account of certain phonological and analogical developments that have occurred diachronically.

2.3.1 THE ROMANIAN CASE SYSTEM

In modern Romanian, the 'nominative-accusative' is distinguished morphologically from the 'genitive-dative' in the singular of most feminine nominals.[79] The nominative-accusative form in nominals is used to encode the subject and direct object.[80] The main function of the genitive-dative is indirect object marking.[81] Some nouns also distinguish a vocative, though it will not be considered here.[82] Romanian has two number values (singular and plural) and grammars list either two or three grammatical genders, where some scholars count only the feminine and masculine,[83] and others include a third, 'neuter' gender.[84] This latter is referred to throughout as *genus alternans* after Igartua (2006), and the view that "nouns showing *genus alternans* are not a class defined by the agreement behaviour of associated words, but a

[78] Determiners and pronouns more faithfully distinguish different case values. They are not considered here, although the enclitic definite article is briefly discussed in §2.3.1.
[79] Daniliuc and Daniliuc (2000:xiii), Gönczöl (2008:29), and Bateman and Polinsky (2010:65) list five case values: nominative, accusative, genitive, dative, and vocative. Whilst some of these distinctions are found in pronominal systems, identifying five case values does not accurately reflect the synchronic state of nouns and adjectives in the language.
[80] Only the first and second person personal pronouns distinguish nominative and accusative functions morphologically. Subject and direct object functions in (indefinite) nominals are usually differentiated not by morphology, but by word order, context, and/or clitic pronouns. Pană Dindelegan (2013c:100).
[81] Maiden (2016a:101).
[82] Manu Magda (2013:590-596) and Maiden (2016a:103) discuss the forms and uses of the Romanian vocative.
[83] Acquaviva (2008:135-40), Gönczöl (2008), Nedelcu (2013a), and Pană Dindelegan (2013a).
[84] Pătruţ (1956), Posner (1996), Bateman and Polinsky (2010), and Maiden (2016a and d).

class of the agreement behaviour of whose associated words is dictated by inflexional morphology" is adopted.[85] That is, these nouns take masculine agreement in the singular but take feminine agreement in the plural, and are thus nouns which exhibit 'alternating gender' (as opposed to 'neither' masculine or feminine gender, as the term 'neuter' implies).

Romanian exhibits a great deal of fusional structure in the inflexion of its nominals, including morphophonological alternations of the lexical stem "which co-occur with the affixal morphosyntactic markers [...]; [these alternations] are occasionally the sole markers (although they co-occur *orthographically* with an affixal marker)".[86] Four sub-classes (or inflexion classes) of two-form feminine nominals are identified by Nedelcu (2013b:275):

	{FATĂ} 'girl'		**{STEA} 'star'**		**{COADĂ} 'tail'**		**{FAMILIE} 'family'**	
	Sing.	**Pl.**	**Sing.**	**Pl.**	**Sing.**	**Pl.**	**Sing.**	**Pl.**
Nom.-acc.	*fată*	*fete*	*stea*	*stele*	*coadă*	*cozi*	*familie*	*familii*
Gen.-dat.	*fete*	*fete*	*stele*	*stele*	*cozi*	*cozi*	*familii*	*familii*

Table 2.15: Selected Romanian lexemes fully declined (non-definite)

Declensions in Romanian are traditionally differentiated according to suffixal alternations,[87] despite frequent concurrent occurrences of stem allomorphy, not only in feminine nominals but in masculine and *genus alternans* lexemes, too. Including stem alternations exhibited by feminine forms, Loporcaro identifies the thirteen inflexion classes given in table 2.16.

[85] Maiden (2016d:119). Since *genus alternans* forms do not distinguish case, the decision not to consider this a separate gender does not affect the arguments put forward later.
[86] Stan and Pologea (2013:607).
[87] Nedelcu (2013b), Giurgea (2013:839-841).

	Romanian feminine nouns		IC	Gloss
a.	*student-ă*	*student-e*	A-*ă*/A-*e*	student (female)
b.	*băutur-ă*	*băutur-i*	A-*ă*/A-*i*	drink
c.	*lips-ă*	*lips-uri*	A-*ă*/A-*uri*	defect, absence
d.	*librări-e*	*librări-i*	A-*e*/A-i	bookshop
e.	*vrem-e*	*vrem-uri*	A-*e*/A-*uri*	time, weather
f.	*zi*	*zi-le*	A-*Ø*/A-*le*	day
g.	*cut-e*	*cut-e*	A-*e*/A-*e*	whetstone
h.	*fat-ă*	*fet-e*	A-*ă*/B-*e*	girl
i.	*sal-ă*	*săl-i*	A-*ă*/B-*i*	room, hall
j.	*brânz-ă*	*brânzet-uri*	A-*ă*/B-*uri*	cheese[88]
k.	*floar-e*	*flor-i*	A-*e*/B-*i*	flower
l.	*sar-e*	*săr-uri*	A-*e*/B-*uri*	salt
m.	*stea*	*ste-le*	A-*Ø*/-B-*le*	star

Table 2.16: Inflexional patterns exhibited by Romanian feminine nouns[89]

Very often the shape of one inflected form in the paradigm of two-form feminine nominals is not predictable from that of the other.[90] The forms in the inflexional paradigm instead must be learned to ensure the correct form is produced. Despite the many different types of stem and suffixal alternations, the vast majority of feminine nominals exhibit the same 'morphomic' distribution of forms within the paradigm:[91]

	Singular	Plural
Nominative-accusative	FORM1	FORM2
Genitive-dative	FORM2	FORM2

Table 2.17: Patterns of syncretism in Romanian feminine nominals

The distribution of [FORM2] cannot be analysed as reflecting to a 'natural class' of morphosyntactic features: it is not consistently aligned solely with 'plural' (as it is found in the genitive-dative singular cell), nor with 'genitive-dative' (as it is found in the nominative-

[88] This is the sole member of this class. Martin Maiden (p.c.).
[89] Adapted from Loporcaro (2017:96). Similar patterns are found in adjectives.
[90] "Feminine [nominative-accusative] singulars in *-ă* alternate with forms in *-e* or *-i*, but it is impossible to predict which: *-e* tends to predominate with animate reference (*românc*ă*-românce* 'Romanian woman', cf. *ţigancă-ţigănci* 'gypsy woman')." Maiden (2016a:101). The plural forms of masculine and *genus alternans* nominals are similarly difficult to predict given knowledge of the singular form (and vice versa). One exception is that, where the singular of a masculine animate noun ends in a consonant *-u*, its plural will end in *-i*. Maiden (2016a).
[91] The term 'morphome' (from Aronoff 1994) here and throughout refers to "an abstract function systematically mapping a single form on to a heterogeneous assembly of grammatical functions." Maiden (2015:33).

accusative plural cell). Nevertheless, the pattern offered in table 2.17 is quite informative about the distribution of stem shapes in the paradigm: where the plural form is known, in the vast majority of instances, a speaker can determine the form of the genitive-dative singular with absolute certainty, since the distribution of forms within the inflexional paradigm of feminine nominals is almost entirely regular across the entire language.

There are two groups of lexemes which exhibit exceptional paradigmatic patterns. In the first place there are a handful of three-form mass nouns, which, instead of exhibiting syncretism of the genitive-dative singular with both cases in the plural, have a third form in *-uri* which is phonologically distinct from the genitive-dative singular (as well as the nominative-accusative singular).[92] Second, there are a few instances of invariant feminine lexemes, including mass nouns such as the singulars {LINTE} 'lentils' and {FASOLE} 'beans', but also count nouns such as {IESLE} 'crib', {MERINDE} 'snack, packed lunch', and {VINTRE} 'belly'.

Masculine and *genus alternans* nominals are invariant for case and only distinguish number morphologically (although some lexemes are formally invariant). As for feminine nominals, in masculine and *genus alternans* lexemes, the plural form is almost never predictable from that of the singular; the distribution of these forms within the inflexional paradigm differs from the distribution of feminine inflected forms, but is consistent according to gender.

Throughout the language, morphological case is more faithfully distinguished by determiners than nominals. For instance, the definite article cumulatively expresses case, number, and gender and can occur with nominals of all genders:[93]

[92] Maiden (2015:44f.) gives an almost exhaustive list of the three-form feminine lexemes.
[93] Since the definite article in Romanian is often analysed as a suffix (rather than a clitic) (e.g., Dragomirescu and Nicolae 2011, Giurgea 2013, and Ledgeway 2017), it is included in the discussion offered here as a

	{FATĂ} - girl			{ELEV} - pupil	
Feminine	**Singular**	**Plural**	**Masculine**	**Singular**	**Plural**
Nom.-acc.	*fata*	*fetele*	**Nom.-acc.**	*elevul*[94]	*elevii*
Gen.-dat.	*fetei*	*fetelor*	**Gen.-dat.**	*elevului*	*elevilor*

Table 2.18: Romanian definite declensions[95]

The definite article always occurs with the form of the nominal which expresses the relevant case and/or number functions.[96] So, for instance, the feminine genitive-dative singular definite article *-i* occurs with the genitive-dative singular form *fete*. Case and number values are therefore cumulatively expressed by an inflected form of the nominal in addition to the relevant shape of the enclitic definite article (which also distinguishes gender in all forms but the genitive-dative plural).

2.3.2 THE DEVELOPMENT OF THE ROMANIAN CASE SYSTEM - PHONOLOGY

The Romanian two-case system is traditionally thought to have been inherited from Latin. In third declension feminine nominals, it is possible to explain the Romanian nominative-accusative/genitive-dative case distinction in the singular as (largely) the result of regular phonological developments. It is argued that the Romanian genitive-dative singular form continues the original Latin dative singular because, in third declension feminine nouns, the nominative/accusative plural in *-ēs* could have merged phonologically with the same form as the dative/ablative plural form in *-īs* more easily than with the genitive plural in *-um* (as seen in table 2.19).[97]

---

cumulative marker of definiteness *and* case(/number/gender). I will continue to call this form the '(enclitic) definite article'.

[94] The nominative-accusative definite article has two allomorphs: *-(u)l*, which occurs with nominals endings in a consonant, and *-le*, which is found with masculine nominals ending in *-e*. Stan (2013:286) and Maiden (2016a:107).

[95] Table 2.18 adapted from Gönczöl (2008:36f.) and Stan (2013:286-291). The *genus alternans* definite articles are the same as those found in masculine lexemes in the singular and the same as feminine lexemes in the plural.

[96] The definite article in modern Romanian occurs with common nouns, proper names, and prenominal adjectives. Stan (2013:286).

[97] Maiden (2015:38), Meyer-Lübke (1985:11). Furthermore, this may offer some insight into why continuants of the Latin second declension do not distinguish case in the singular, "for all the forms of the Latin second

	Latin Singular	Romanian Singular	Latin Plural	Romanian Plural
Nominative	(LĒX)	*lege*	LĒGĒS	*legi*
Accusative	LĒGE(M)		LĒGĒS	
Genitive	LĒGIS	*legi*[98]	LĒGUM	(**legu*)
Dative	LĒGĪ		LĒGIBUS > **lēgīs*[99]	*legi*

Table 2.19: Romanian {LEGE} 'law' from Latin third declension (feminine) {LĒX} 'law'[100]

The morphological case distinction found in the paradigm of Romanian lexemes with a nominative-accusative singular form ending in *-ă* is also considered to be the result of regular phonological developments affecting first declension Latin nominals. In the first place, the Latin nominative and accusative forms fell together phonologically in the singular,[101] leading to a form in *-ă*; the Romanian genitive/dative singular developed from the Latin genitive/dative singular form, where *-ae* > *-e*.

Upon closer consideration, there are a couple of weaknesses associated with the analysis given above. In the first place, in the plural of lexemes inherited from the Latin first declension, it is not clear why case distinctions should not be continued, where, according to regular phonological developments, no neutralization should have occurred. Additionally, as Maiden points out, if the Romanian feminine singular genitive-dative desinence in *-e* does continue Latin *-ae*, "one would expect the resultant *-e* to have triggered regular palatalization of root-final velar consonants",[102] of which development there appear to be no examples in Aromanian or Istro-Romanian. Evidence such as this suggests that the development of *-e* occurred *later* than the palatalization change, which leads linguists such as Miklosich

declension singular, *other than the genitive*, would have converged by regular sound change". Maiden (2015:37f.). Emphasis added.

[98] Maiden (1996:168f.) discusses the development *-is* > **-es* > *-i*.

[99] The form LĒGIBUS did not become **lēgīs* by regular phonological change, but was rather modelled analogically on the first declension feminine dative/ablative plural form.

[100] Adapted from Maiden (2015).

[101] A development discussed in §2.1.2.

[102] 2015:40.

(1882:15f.), Philippide (2011), and Maiden (2015) to argue that the Latin case system was actually completely lost from proto-Daco-Romance and that the case distinction in feminine singulars in Romanian is a subsequent endogenous development.[103] The distinction, it is argued,

> "is the regular phonological effect of combining the Daco-Romance feminine singular ending -ə with a following enclitic [genitive-dative] form of the feminine (originally *-eị), which allegedly caused fronting of the -ə to -e, due to an assimilatory fronting of schwa before a following /e/".[104]

Since no sound change affected the masculine (or *genus alternans*) forms in this way, these lexemes remain invariant for case.

The total loss of case distinctions particularly in masculine lexemes inherited from the Latin third declension cannot be attributed solely to phonological changes. Moreover, the analysis offered in table 2.19 reconstructs an unetymological dative/ablative form to account for the syncretism of the nominative-accusative and genitive-dative in the plural of Romanian feminine and masculine nominals. In the data set interrogated by Gaeng, howeover, there is only one instance (although it is, notably, from Dacia) in which the dative plural *-ibus* in a third declension lexeme is 'replaced' with *-īs*: *posuerunt ... patris benemerentis* where *patribus benemerentibus* would be expected.[105] But there is little other independent evidence to support the suggestion that the dative/ablative plural and nominative/accusative plural forms of etymologically third declension lexemes 'fell together' phonologically in Romanian. Thus, it is concluded here, as it was in reference to Latin (§2.1.3) and old Gallo-Romance (§2.2.3) that phonetic erosion alone cannot account for the diachronic developments exhibited by the Romanian case system.

[103] Some sub-Danubian dialects feminine singular nominals do not inflect for case, perhaps lending further support to the hypothesis that cases were a secondary development in the history of Romanian.
[104] Maiden (2015:38f.).
[105] 1984:81.

2.3.3 THE DEVELOPMENT OF THE ROMANIAN CASE SYSTEM - ANALOGY

If the Romanian case system is actually a language-internal innovation (as outlined in the previous section), then the spreading of case-marking throughout the language to the vast majority of feminine nominals may be analysed as an analogical development. It is argued by Maiden (2015) that the alternation originated in lexemes with a nominative-accusative singular form in *-ă*. Since almost all nominals with this structure in the nominative singular are feminine,[106] morphological case-marking may have become associated with feminine gender, and thereafter spread to other feminine nominals. On the model of *fată*-type nominals, the form of the plural was extended to distinguish genitive-dative singular functions, replacing the etymological form (which would have been identical to the nominative-accusative singular). Some evidence for this position can be found in lexemes which lack a plural, and which are thereore invariant for case in the singular, since "the creation of a distinct singular [genitive-dative form] is blocked."[107]

Further support for this view may come from the behaviour of lexemes with a nominative-accusative singular form in *-ă* compared to that of lexemes with a nominative-accusative singular ending in anything other than *-ă*. That is: invariance is only ever found in forms with a nominative-accusative singular ending in something other than *-ă*. If the case distinction was one inherited from Latin, according to regular phonological developments those feminine lexemes inherited from the Latin third declension (or, for that matter, in masculine nominals inherited from the Latin third declension) should have maintained a morphological difference between the 'nominative-accusative' and the 'genitive-dative'. If, howeover, the original morphological distinction is the result of a phonological development affecting only those

[106] There are a very few exceptions to this generalization: *tată* 'father' and *popă* 'priest' exhibit what seems to be feminine morphology and yet are masculine. Intriguingly, the feminine form of the enclitic definite article can be found with both of these forms and, in the definite declension, these lexemes can exhibit a distribution of forms like that found in feminine nominals, i.e., distinct stem shapes in the nominative-accusative singular and genitive-dative singular forms. Nedelcu (2013c:277).
[107] Maiden (2015:42).

forms ending in *-ă*, it would have been restricted to those lexemes which historically descend from Latin first declension nominals. If the morphophonological distinction between the nominative-accusative and genitive-dative forms then analogically spread through the language, we are likely to find: a) instances in which the extension of a genitive-dative singular form distinct from the nominative-accusative singular was 'blocked' (because of the lack of a distinct plural form), and b) potentially, incompleteness of the analogical extension *only* in those forms with a nominative-accusative singular form ending in anything other than *-ă*. The fact that both of these expectations are met may provide some further support for the endogenous innovation theory of the Romanian case system.

The morphomic distribution of forms in the feminine paradigm appears to be recognized and utilized by speakers: "throughout the observable history of the language any morphological change affecting any one of the plural or [genitive-dative] singular cells in Romanian feminines, whatever its nature, has affected all of the others in the same way."[108] For instance, where in modern Romanian the final *-e* found in a genitive-dative singular form has been replaced by *-i*, it is not just the plural form which is affected by this development, but also the genitive-dative singular form. For instance: {GURĂ} 'mouth' exhibited an older genitive/dative singular (and a plural) form *gure*, but final *-e* has been replaced by *-i*, giving *guri*. Similarly, {ROATĂ} 'wheel' exhibits two genitive/dative singular (and two plural) forms: *roate* and *roţi.* These two suffixes, *-e* and *-i*, are in competition in the modern language.[109] The form of the genitive-dative singular in these instances is not affected by phonology, but is analogically altered to be formally identical to the shape of the plural. The distribution of feminine forms thus appears to be recognized by speakers and exploited in language use. This will be discussed in further detail in §6.3.

[108] Maiden (2015:43). Three-form feminine lexemes are rare counter-examples to this generalization.
[109] Nedelcu (2013b:275). Examples from Maiden (2015:43).

2.4 CLOSING REMARKS

The development of the Latin nominal case system is, from a descriptive point of view, simple: from a five-case system in the Classical language, the morphological distinctions were gradually lost. Only one modern Romance language, Romanian, distinguishes case morphologically in nominals, and even then the difference between nominative-accusative and genitive-dative is only found in feminine singular lexemes (nor is it clear that this distinction actually directly continues case distinctions found in Latin). The two-case 'system' found in old French and old Occitan nominals has been completely lost from the modern varieties.

Latin and the old Gallo-Romance varieties may safely be described as 'case-reducing'. Romanian, on the other hand, is a little more difficult to analyse. If the two-case system is instead an endogenous innovation, should Romanian be described as 'case-increasing'? Although I lean towards Maiden's (2015) analysis of the Romanian case system, I choose to describe the language as 'case-steady': the two-value system is found since the earliest attestations of the language (although admittedly this does not give us a very significant time-depth of evidence from which to draw), and continues to be used and understood to this day. Furthermore, there are no indications, to the best of my knowledge, that there are any new morphological case-forms in development (as there clearly are in Veps, for instance - §3.3.1) to provide further support for an analysis of unambiguously case-increasing. For these reasons, Romanian is here analysed as 'case-steady'.

The intertwined, e.g., phonological, analogical, functional, and syntactic developments outlined in this chapter account for some of the particular developments found in the Italic languages under discussion. These considerations can often provide information about the loss of one or two morphological case distinctions (usually only in a certain inflexion class or

in nominals of a certain gender), although in many instances our analyses have been necessarily speculative as the result of the relative paucity of relevant data. However, even combined, these factors do not satisfactorily account for the total loss of case from the majority of Italic varieties, nor for the supposed innovation of a case distinction in Romanian; it thus remains unclear *why* these case systems have developed as they have.

Later in this study, I will suggest that consideration of the distinct types of morphological structure found in the data presented here and understanding the ways in which these structures are learned and exploited by language-users is essential to understanding why certain languages will be likely to lose, maintain, or propagate case-forms. That is not to say that other factors such phonological, analogical, and syntactic developments, as well as contact with other languages, did *not* affect the case systems of the varieties discussed here, but that consideration of the data from a slightly different perspective may shed further light on why certain developments have occurred in particular languages through time.

3.

THE FINNIC LANGUAGES

3.0 THE FINNIC CASE SYSTEM

It is in many ways more difficult to discuss the development of the Finnic than the Italic case system. Although there are gaps in the attested 'stages' of the transition of Latin to its multiple Romance descendants, Latin does at least provide information about an earlier stage of the old Gallo-Romance and Romanian case systems, which offers insight into certain changes which occurred in the history of these varieties. The earliest attestation of any Finnic language, on the other hand, dates from the first half of the thirteenth century C.E.,[1] and no Finnic variety is extensively written until the sixteenth century Finnish works of Agricola.[2] This represents a much later and much more limited corpus than that offered by Latin.

In order to offer some discussion on the developments which happened before the Finnic varieties were first attested, I will refer to 'proto-Finnic' and to earlier stages of the case system. Although the Finnic languages exhibit a great number of lexical, syntactic and morphological similarities, and some of the reconstructions are offered with a great deal of confidence, it should be remembered that these postulations do not always accurately reflect what was found at an earlier stage of a given language or language family.[3] In addition to historical changes that occurred before the earliest attestations of the Finnic languages, I will discuss more recent and ongoing developments affecting these varieties.

[1] The Birchbark letter no. 292. Haavio (1964:46).
[2] Karlsson (2008:3). The first Livonian literature comes from the nineteenth century (Ernštreits 2007, Grünthal 2015b); the earliest written examples of Veps come from the early 1930s (Blokland and Hasselblatt 2003:128).
[3] Attempts to reconstruct Latin from the modern Romance languages demonstrate this point quite neatly. For instance, comparison of forms for the word 'tooth' in various modern Romance varieties (French *dent*, Spanish *diente*, Italian *dente*, Romanian *dinte*, for instance) lead us to reconstruct a proto-Romance form **dente*; these forms actually descend from the Latin accusative singular form DENTEM (where the final *-m* has been lost without a trace in these forms in modern Romance). Without knowledge of Latin we cannot reconstruct the final nasal, nor do we have sufficient information to determine that this is actually the accusative singular form of the lexeme [DĒNS].

3.1.0 INTRODUCTION TO FINNISH

Finnish has the largest number of speakers of any of the Finnic varieties, with nearly five million native speakers in Finland,[4] and smaller communities in particularly Sweden and Russia. The following sections will offer an introduction to the synchronic Finnish case system (§3.1.1), and thereafter discuss certain developments, including some which are postulated to have occurred since the splitting of proto-Finnic and others which are ongoing in the modern language (§§3.1.2-5).

3.1.1 THE FINNISH CASE SYSTEM

Most Finnish grammars usually count between twelve and seventeen case values (given in table 3.1).[5] Not all nominals have prolative or lative forms; the instructive, comitative, and abessive forms are in restricted usage in the modern language and are most frequently found as fossilized forms; the prolative is productive only where semantics allow. Furthermore, direct object functions are not distinguished by a unique form in nominals.[6] The remaining values in table 3.1 are all productive in Finnish and constitute the 'core' of the inflexional case system. I will omit from the following discussion the 'accusative' (since it is never encoded by a unique form in nominals) and the lative (since it is found only in the paradigms of a few demonstrative pronouns). Although the instructive, comitative, prolative, and abessive are semantically and/or stylistically restricted, the development of these forms will be included in this study.

[4] http://www.stat.fi/tup/suoluk/suoluk_vaesto_en.html. Accessed 11/04/2019, 10:37.
[5] ISK §1221, Hakulinen (1979:97), and Karlsson (2008:119-191) count fifteen, omitting the lative and prolative; the abessive, instructive, and comitative are described as 'adverbial' by Abondolo (2006c:157), who counts twelve. Collinder (1965:54f.) gives the seventeen listed in table 3.1.
[6] In personal pronouns and the interrogative pronoun *kuka* the accusative is distinguished by means of a suffix *-t*.

	{PUHELIN} 'telephone'		{AJATUS} 'thought'	
	Singular	**Plural**	**Singular**	**Plural**
Nominative	*puhelin*	*puhelime-t*	*ajatus*	*ajatukse-t*
Accusative	*puhelin/ puhelime-n*	*puhelime-t/ puhelim-i-a*	*ajatus/ajatukse-n*	*ajatukse-t/ ajatuks-i-a*
Genitive	*puhelime-n*	*puhelim-i-en/ puhelin-ten*	*ajatukse-n*	*ajatus-ten/ ajatuks-i-en*
Partitive	*puhelin-ta*	*puhelim-i-a*	*ajatus-ta*	*ajatuks-i-a*
Essive	*puhelime-na*	*puhelim-i-na*	*ajatukse-na*	*ajatuks-i-na*
Translative	*puhelime-ksi*	*puhelim-i-ksi*	*ajatukse-ksi*	*ajatuks-i-ksi*
Elative	*puhelime-sta*	*puhelim-i-sta*	*ajatukse-sta*	*ajatuks-i-sta*
Inessive	*puhelime-ssa*	*puhelim-i-ssa*	*ajatukse-ssa*	*ajatuks-i-ssa*
Illative	*puhelime-en*	*puhelim-i-in*	*ajatukse-en*	*ajatuks-i-in*
Ablative	*puhelime-lta*	*puhelim-i-lta*	*ajatukse-lta*	*ajatuks-i-lta*
Adessive	*puhelime-lla*	*puhelim-i-lla*	*ajatukse-lla*	*ajatuks-i-lla*
Allative	*puhelime-lle*	*puhelim-i-lle*	*ajatukse-lle*	*ajatuks-i-lle*
Instructive	*puhelim-i-n*		*ajatuks-i-n*	
Comitative	*puhelim-i-ne-*[7]		*ajatuks-i-ne-*	
Abessive	*puhelime-tta*	*puhelim-i-tta*	*ajatukse-tta*	*ajatuks-i-tta*
Lative	-	-	-	-
Prolative	*puhelim-i-tse*		-	

Table 3.1: Finnish {PUHELIN} and {AJATUS} fully declined

The main functions of the grammatical cases are as follows:

- nominative: the form of the subject of a finite verb; the form of the direct object with verbs of necessity as well as in passive and imperative constructions;
- genitive: used both adnominally (to express possession) and adverbally (as the telic direct object, or the subject in constructions expressing necessity); generally the inflected form of the complement with postpositions;
- partitive: encodes the indefinite subject and indefinite object;[8] used also when counting, and as the object in negative sentences; generally the inflected form of the complement with prepositions.

[7] The comitative ending is always followed by a possessive suffix in nouns, though not in adjectives.
[8] The nominative, genitive, and partitive in the singular can all be used as the case-form of the direct object. *In nuce*: in the singular, the genitive form is used for definite, telic objects in affirmative sentences; the partitive form is used for the indefinite or atelic object (including in negative sentences), and the nominative is only used

The main functions of the remaining case-forms are as follows:

	Case-form (singular and plural)	Main meanings
Essive	*-nA*[9]	[a (temporary) state or function]; 'on, at' (temporal, occasionally local)
Translative	*-ksi*	[the end point of movement or change]
Ablative	*-ltA*	'away from'; 'since'
Adessive	*-llA*	'on, at'; 'in, during'; [possession]; [instrument]; [manner]
Allative	*-lle*	'towards, onto, to'
Elative	*-stA*	'out of, from'; 'about'; '(made) out of'
Inessive	*-ssA*	'in'; 'during'
Illative	*-Vn/-seen/-hVn*[10]	'into, to'
Instructive	*-in*	[manner]
Comitative	*-ine-*	'with'
Abessive	*-ttA*	'without'
Prolative	*-itse*	'across'

Table 3.2: The functions of the 'semantic' Finnish cases[11]

In the Finnic languages, as briefly discussed in §1.2.2, inflexion classes are distinguished by allomorphy in the stem. Most grammars identify 'basic' and 'inflexional' stems in Finnish nominal declensions, the basic stem being found in the nominative singular and the inflexional stem being found elsewhere. Sometimes a separate 'plural' stem is also identified, which is distinct from other stem shapes in the paradigm and is found in all plural forms except the nominative (which is instead characterized by the inflexional stem). I identify four different stem shapes and to give them slightly different names to those offered in traditional grammars:[12]

---

as the form of the direct object when the subject is not encoded by a nominative form. In the plural the nominative form is used for definite, telic objects. Karlsson (2008:158-164).

[9] Here and throughout discussion on the Finnish morphological case system, *-A*, *-O*, and *-U* are used to indicate that the vowel obeys the constraints of vowel harmony (described later in this section).

[10] When the illative forms are discussed, *-V* stands for a vowel which is identical to that immediately preceding the inflexional suffix.

[11] Karlsson (2008:119-191), ISK §1221-1266. As discussed in §1.2.2, in the Finnic languages the nominative, genitive, and partitive are described as the 'grammatical' case-forms; the remaining values as the 'semantic' case-forms, a convention that is adopted here.

[12] These names are completely arbitrary.

- the nominative stem: similar to the 'basic' stem in traditional accounts, this is the stem shape found in the nominative singular; it is occasionally the same stem shape as that found in the partitive singular form, and, less frequently, the same shape as the 'common' stem (discussed below);
- the partitive stem: occasionally, the partitive singular is characterized by a stem shape distinct from that of the nominative and the common stems (discussed below);
- the common stem: generally found in all singular forms apart from the nominative (frequently) and the partitive (occasionally); also the stem shape found in the nominative plural form;
- the plural stem: the shape of the stem found in all plural forms apart from the nominative.

The common and plural stems may occasionally have two allomorphs, referred to as the 'weak' and the 'strong' grades of the stem. The alternation of these forms has to do with qualitative or quantitative gradational differences which affect the consonants *p*, *t*, and *k* and their geminate counterparts. The alternations of quantitative consonant gradation (strong ~ weak) are as follows:

- *pp* ~ *p*
- *tt* ~ *t*
- *kk* ~ *k*

The alternations of qualitative consonant gradation (strong ~ weak) are as follows:[13]

[13] Qualitative consonant grade alternations are more complicated than is presented here when the stops *p*, *t*, and *k* are in a consonant cluster. When preceded by a nasal, for instance, *p* and *t* assimilate retrogressively; *nk* becomes *ng*; when preceded by *l* or *r*, *t* retrogressively assimilates. Karlsson (2008:50-52).

- *p ~ v*
- *t ~ d*
- *k ~ v/Ø*[14]

The distribution of weak and strong forms in Finnish can usually be expressed in purely phonological terms, where in most nominal-types the weak stem occurs in a closed syllable and the strong stem is found in an open syllable. So, for instance, the nominal {SUKKA} 'sock' given in table 3.3 below has two common stems, one with long *-kk-* and the other with the short counterpart *-k-* (i.e., *sukka- ~ suka-*). Whenever the syllable following the medial consonant *-kk-* is closed, the long consonant degeminates.

There are, however, rare instances of an unexpected grade. For instance, the nominative singular forms of *liike*-type nominals are characterized by the weak stem, despite exhibiting a strong stem in all other forms in the paradigm (apart from the partitive singular),[15] e.g., genitive singular *liikkeen* 'of the business'. Examples such as these indicate that consonant gradation alternations have morphologized in the modern language and are no longer phonologically-conditioned alternations.[16] Nevertheless, for most lexemes, the distribution of different stem grades generally follows the same rules: strong stem shapes are found in open syllables; weak stem shapes are found in closed syllables.

Karlsson counts four major declensions of Finnish nominals in which the nominative stem is distinct from that of the common stem, which are further split into fourteen subtypes

[14] Additionally *k ~ g* (phonetically [ŋ]) but only when preceded by an *-n-*, e.g. *kenkä* 'shoe', nominative plural *kengät* [kɛŋŋæt]; the alternation *k ~ v* occurs between labials and *k ~ Ø* elsewhere.

[15] Historically, the nominative singular of *liike*-type nominals would have ended in **-k* (Hakulinen 1979:46 Abondolo 2006c:162f.), which closed the syllable. This final **-k* has left a trace in Finnish whereby following consonants are lengthened (as in the partitive singular, the shape of which is the result of an assimilation **-kt- > -tt-*). Hakulinen (1979:53).

[16] Further evidence for this analysis comes from recent loanwords. For instance, the genitive singular form of *auto* 'car' is *auton*, where ***audon* would be expected.

according to different types of stem allomorphy; there is a further, fifth declension in which the nominative, partitive, and common stems are all the same shape:[17]

	Nominative stem	**Partitive stem**	**Common stem(s)**[18]	**Plural stem**[19]	**Gloss**
Nominals with a nominative stem in *-i*	*tunti*	*tunti-, tunni-*		*tunte-i-, tunne-i-*	hour; lesson
	lehti	*lehte-, lehde-*		*leht-i-, lehd-i-*	leaf; newspaper
	kieli	*kiel-*	*kiele-*	*kiel-i-*	language; tongue
	vesi	*vet-*	*vete-, vede-*	*ves-i-*	water
Nominals with a common stem in *-ee*[20]	*liike*	*liiket-*	*liikkee-*	*liikke-i-*	movement; business
Nominals with a nominative stem in *-A*	*sukka*	*sukka-, suka-*		*suk-i-, sukk-i-*	sock
	omena	*omena-*		*omeno-i-*	apple
Nominals with a nominative stem ending in a consonant	*ihminen*	*ihmis-*	*ihmise-*	*ihmis-i-*	person
	ajatus(-)		*ajatukse-*	*ajatuks-i-*	thought
	rikas(-)		*rikkaa-*	*rikka-i-*	rich
	uutuus	*uutuut-*	*uutuute-, uutuude-*	*uutuuks-i-*	novelty
	avain(-)		*avaime-*	*avaim-i-*	key
	onneton(-)		*onnettoma-*	*onnettom-i-*	unhappy
	jäsen(-)		*jäsene-*	*jäsen-i-*	member
	lyhyt(-)		*lyhye-*	*lyhy-i-*	short
Nominals with identical nominative/ common stem	*katu-, kadu-*			*katu-i-, kadu-i-*	street
	maa[21]			*ma-i*	land, countryside
	nalle			*nalle-i-*	teddy

Table 3.3: Finnish declensions[22]

[17] 2008:61-82.
[18] Where two stems are listed, the strong stem precedes the weak.
[19] The form *-i-* included in these forms is generally analysed as the plural marker.
[20] Karlsson (2008:68) refers to these as 'nominals with a basic form in *-e*', but, since there are nominals with a basic form ending in *-e* such as *nalle* 'teddy' (the last nominal type listed in table 3.3) which decline in a slightly different way to nominals such as *liike*, it is changed here to 'nominals with a common stem in *-ee*'.
[21] *maa*-type nominals include monosyllabic nominals ending in a long vowel or a diphthong and differ from other nominal types in that they exhibit an illative suffix *-hVn*, a form unique to this declension class.

Unlike lexical stems, many morphological case suffixes in the Finnish case system are invariant across declension and number. There are two exceptions where, across inflexion classes, certain case values are discriminated by distinct desinences. The first of these exceptions is that of vowel harmony where, generally in the modern language, "[i]f the stem [of a lexeme] contains one or more of the vowels **u**, **o**, **a**, the ending also has to have a back vowel (**u**, **o**, **a**). If the stem has no back vowels, the ending has to have a front vowel (**y**, **ö**, **ä**)."[23] The vowels *i* and *e* are 'vowel harmony neutral' and can occur with back or front vowels. Synchronically, the rules of vowel harmony hold across the vast majority of native forms,[24] although, since loanwords frequently have both back and front vowels, it is not always clear to language-users whether inflexional (and derivational) material should have back or front vowels.[25]

In addition to vowel harmony differences across lexemes, there are a handful of examples of case values which are characterized by distinct desinences across declension, across number, or sometimes both. The nominative singular form, for instance, is always characterized by a lack of overt suffix, whilst the nominative plural marker is usually analysed as *-t*. The genitive in the singular is always distinguished by *-n*, whilst in the plural there are five genitive exponents. The illative (singular and plural), partitive (singular and plural), and genitive plural suffixes are the only case-forms which vary formally across inflexion class.

---

[22] There are many *liike-*, *katu-*, and *sukka*-type nominals which do not exhibit different stem grades. Whilst the overview offered in table 3.3 tries to be representative of much of the inflexional variation found in Finnish, it should be noted that "[t]here is no consensus on how many inflectional classes there are for nominals." Karlsson (2006:476).

[23] Karlsson (2008:20).

[24] There are just two examples of unexpected vowel qualities in the inflexional case endings: since there are no back vowel in the stems *mer-* 'sea' and *ver-* 'blood', the partitive singular forms ***mertä* and ***vertä* are expected; instead, back vowels are found: *merta*, *verta*. This exception occurs only in the partitive singular forms, where other inflexional suffixes have the expected front vowels, e.g., adessive singular *merellä*, inessive singular *veressä*. ISK §15.

[25] Maamies (2015), ISK §17.

Case	Case-form	Nominal Type[26]
Illative Singular	*-Vn*	*lehti, kieli, vesi, sukka, omena, ihminen, ajatus, uutuus, avain, onneton, jäsen, lyhyt, katu, tunti, nalle, valo*
	-hVn	*maa*
	-seen	*liike, rikas*
Illative Plural	*-i-in*	*lehti, kieli, vesi, sukka, omena$_1$, ihminen, ajatus, uutuus, avain, onneton, jäsen, lyhyt, katu*
	-i-hin	*tunti, maa, omena$_2$, nalle*
	-i-siin/-i-hin[27]	*liike, rikas*
Partitive Singular	*-A*	*tunti, lehti, sukka, omena, katu, nalle, valo*
	-tA	*kieli, vesi, ihminen, ajatus, rikas, uutuus, avain, onneton, jäsen, lyhyt, maa*
	-ttA	*liike*
Partitive Plural	*-i-A*[28]	*tunti, lehti, kieli, vesi, sukka, omena$_1$, ihminen, ajatus, uutuus, avain, onneton, jäsen, katu, nalle, valo*
	-i-tA	*liike, rikas, lyhyt, maa, omena$_2$*
Genitive Plural[29]	*-i-en*	*sukka$_1$, omena$_1$, kieli$_1$, vesi$_1$, ihminen$_1$, uutuus, ajatus$_1$, avain$_1$, onneton$_1$, jäsen$_1$, tunti, katu, nalle$_1$, valo*
	-i-n	*sukka$_2$, omena$_3$,*[30] *nalle$_2$*
	-i-den[31]	*liike$_1$, rikas$_1$, lyhyt$_1$, maa$_1$, omena$_2$*
	-(i)-tten	*liike$_2$, rikas$_2$, lyhyt$_2$, maa$_2$, vesi$_2$*[32]
	-ten	*ihminen$_2$, ajatus$_2$, avain$_2$, onneton$_2$, jäsen$_2$, kieli$_2$*

Table 3.4: Distribution of case exponents in Finnish

Although not usually included in Finnish grammars, two further case values may be identified in the modern language, although they are lexically or diatopically restricted in their distribution. First, a new comitative form in *-kAA* is found in certain southern dialects, although only in pronouns and animate nouns. It does not licence concord; instead, adjectives that occur in a noun phrase inflected in the comitative are inflected in the genitive:[33]

[26] Some nominal types appear twice because they can take more than one exponent for a given case(/number) value. These are marked with a subscript number for clarity.
[27] The exponent *-hin* is rare. ISK §94, Hakanen (2006:28).
[28] Between vowels, plural *-i-* is written *-j-*.
[29] The endings *-den*, *-tten*, and *-en* attach to the nominative stem; *-ten* and *-in* occur with the plural stem. ISK §83.
[30] In fact, in combination with different stem shapes and distinct genitive plural exponents, {OMENA} 'apple' has five genitive plural forms: *omenien, omenoiden, omenoitten, omenojen*, and *omenain*. Hakanen (2006:23).
[31] Nominal types with a genitive plural in *-i-den* can usually also exhibit a form ending in *-i-tten.*
[32] The alternative genitive plural form *vetten* is irregular and very rare.
[33] In this way, the element *-kAA* behaves very like the Estonian comitative *-ga* which has developed from a cognate form of the Finnish postposition *kanssa* 'with'. Intriguingly, it is those Finnish dialects spoken most closely to Estonia in which the new comitative suffix is developing.

3.1.a *kilti-n* *poja-n-kaa*

kind-GEN.SG. boy-GEN.SG.-COM.

'with the kind boy'

This suffix has developed from the postposition *kanssa* 'with'. It still exhibits some postposition-like behaviour and has not yet fully grammaticalized. It will be discussed further in §3.1.2.

Second, the excessive case is in even more restricted usage than the comitative. It is found only in particular dialects and is not used in the *kirjakieli* (the written language). Its earliest origins can be traced to the 1800s where it first appeared in Finnish grammars written by linguists in the east, though later the excessive form has spread throughout the south and southwest of Finland.[34] Where it occurs with adverbials and postpositions, the excessive usually encodes the meaning 'away from':

3.1.b *talo-n* *taka-nta*

house-GEN.SG. behind-EXC.SG.

'from behind the house'

However, there are also examples of its being used particularly with words denoting jobs to encode the meaning 'to stop being':

3.1.c *läksi-n* *pappila-sta* *palvelija-nta*

leave-1SG vicarage-ELA.SG. servant-EXC.SG.

I left the vicarage, no longer a servant (lit. 'from being a servant')

[34] Särkkä (1969:50f.).

Despite its not being considered a 'fully-fledged' case in modern Finnish, the excessive presents an example of a case-form in development (or having been in development at an earlier stage of the language), and will therefore be discussed in further detail in §3.1.3.

3.1.2 THE DEVELOPMENT OF THE FINNISH CASE SYSTEM - PHONOLOGY

Finnish is often considered the most 'conservative' Finnic variety, since it is thought to have preserved many phonological, morphological, and syntactic aspects reconstructed for proto-Finnic. For instance, the total loss of word-final *-n* in Livonian and Estonian means that the genitive singular is no longer characterized by an invariant suffixal marker. In Finnish, on the other hand, final *-n* is generally retained, although in more rapid speech it can be progressively assimilated or even dropped.[35] In those inflexion classes where there is no stem allomorphy in the singular, the opacification of *-n* can lead to the syncretism of the nominative singular and genitive singular forms. In most nominal types, however, these forms are characterized by different stem shapes which maintain a morphophonological distinction between the nominative and genitive case values.

There is also an ongoing loss of word-final short vowels *-i* and *-A* in Finnish,[36] which is affecting the adessive, ablative, inessive, translative, and essive case-forms. Additionally, the shortening of a final long consonant affects the inessive, allative, adessive, and abessive forms. Thus, a form such as *meressä* 'in the sea' tends to be pronounced *meres* in quick speech. Despite these changes, there appear to be few resulting instances of syncretism (quite different from the results of phonetic erosion outlined in Latin in §2.1.2). For instance, although the inessive suffix is undergoing a change *-ssA* > *-s*, the elative is losing its final short vowel *-stA* > *-st*, and the final *-n* of the illative can be opacified depending on the first sound of a word that follows (as discussed above), distinct forms are maintained in the

[35] For instance, *Teemun pallo* 'Teemu's ball' is often pronounced *Teemum pallo*; *sateen jälkeen* 'after the rain' as *satee(j) jälkeen.* Karlsson (2008:350f.).
[36] Karlsson (2008:347).

internal local case system. Similar observations hold for the external local case-forms, where loss of final short *-A* have affected the ablative and adessive forms (the long final consonant of the adessive has also subsequently shortened), but has not led to the opacification of the morphological distinction of any of the external local functions.[37]

One instance in which phonetic erosion is leading to the loss of morphological distinctions in the case system is that of the developing syncretism of the illative singular (following the opacification of final *-n*) and the partitive singular, e.g., illative singular *kylään* 'into the village' > *kylää*, partitive singular *kylää*. But this syncretism, like that of the nominative singular and genitive singular, is restricted to particular inflexion classes.[38]

Phonetic erosion does not seem to be leading to the loss of case distinctions in modern Finnish, yet the comitative, instructive, prolative, and abessive are falling out of use. These (frequently fossilized) forms are more resistant to the phonological changes described above (i.e., assimilation of *-n*, loss of final short *-A* and *-i*) than the more productive semantic case-forms. For instance, in a fossilized instructive phrase such as *palja-i-n pä-i-n* 'with a bare head', the first instructive ending *-i-n* may assimilate to the following *p-* (i.e., *paljaim pain*), but the second *-n* is pronounced in full.[39] The comitative exponent *-(i)ne-* is very rarely found at the end of the inflected word (since a noun inflected in the comitative takes a necessary possessive suffix); even where it does occur word-finally, this form ends in an *-e* which is not lost from speech. Similarly, the prolative form ending in *-e* is not lost. Finally, the abessive where it occurs in fossilized phrases does not lose its final *-A* and, even if it did undergo

[37] This is not true of all Finnic varieties. In Karelian, for instance, the 'externative' encodes both allative and adessive functions and is characterized by a form *-l* following the loss of short final *-A* and *-e* and the subsequent shortening of the long *-ll*. Vahtia (1938:43f.), Makarov (1963:79), Zajkov (1999:12, 44f.).

[38] A similar 'falling together' of the partitive and illative in the singular is found in Estonian, Votic, and Livonian. However, each of these varieties has developed innovative means of disambiguating these two values morphologically. For instance, in Estonian the illative allomorph *-sse* (Ehala 2009:35), and in the Central and Jõgõperä varieties of Votic the form *-sõ* (Rozhanskiy and Markus 2015:31f.), can be used to disambiguate the illative singular from the partitive singular. The illative singular exponent *-z* in Livonian can occur optionally where the illative singular form is identical to that of the genitive singular. Ainsworth (2019b:19).

[39] My thanks to Ulla-Maija Forsberg (p.c.) for this insight.

similar developments as, for instance, the adessive and inessive (loss of short final *-A*, subsequent de-gemination of final long consonant), the abessive form would still in most lexemes have been distinct from the remaining forms in the paradigm.[40]

Phonetic erosion, therefore, cannot be the sole reason for the gradual loss of these case-forms and their replacement with adpositional phrases or other case-markers:

abessive: *ilman raha-a*

without money-PART.SG. 'without money'

comitative: *vaimo-n kanssa*

wife-COM.SG. with 'with [the, his etc.] wife'

instructive: *(koristeltu) jalokiv-i-llä*

(decorated) jewel-PL.-ADE. 'decorated with jewels'

prolative: *(matkusta-n) maa-ta pitkin*

(travel-1SG) land-PART.SG. along 'I travel by land'

In reference to Latin in §2.1.2, it was suggested that adpositions were increasingly used to differentiate those functions no longer distinguished morphologically. However, the Finnish prolative, comitative, instructive, and abessive are still, in the modern language, phonologically distinct from other inflected forms. The preference for adpositional marking of these functions cannot, therefore, be analysed as the result of opacifying phonological developments in the case system.

[40] In those lexeme types with a partitive exponent *-tA* the formal distinction would have been lost. Notably, though, *liike*-type nominals distinguish partitive and abessive in the singular by means of the stem shape: *liikettä* partitive singular, but *liikkeettä* abessive singular, thus this distinction remains even following the loss of final *-A* and degimination of final long *-tt*.

Phonological developments do, however, play a role in the introduction of novel case suffixes. The new comitative form in *-kAA*, for instance, originates from a phonetically-eroded postposition *kanssa*. As a result of regular phonological developments (discussed above), the final *-a* has been lost, and the final long *-ss-* has degeminated. Moreover, unlike the original postposition, the form *-kAA* can exhibit vowel harmony (although many speakers prefer back vowels): *isänkää* 'with father' (optionally also *isänkaa*), but *munkaa* 'with me' (written language *minun kanssa*). This may be analysed partially as a phonological development. In Finnish, vowel harmony is usually restricted to the morphological word.[41] However, in certain instances, adjacent forms within the speech stream can exhibit vowel harmony alternations. For instance, in fast speech we may find examples such as *ma tulen* or *ma tuun* (rather than the expected *minä tulen*) for 'I come, I am coming', where the front vowel of the first person singular pronoun has changed to its back equivalent when the following verb (and stressed syllable) exhibits a back vowel.[42] The vowel harmony of the form *-kAA* is retrogressive rather than progressive, but the frontness or backness of this final long vowel, particularly when unstressed, may be influenced by the vowel sounds of the constituent with which it is most closely associated semantically and prosodically.

However, phonology can only take us so far (without appealing to *ad hoc* solutions that have little to no generality through the language). The loss of the final *-ns* of the eroded postpositional form **kans*, for example, is not strictly phonological (although the lengthening of the *-a-* might be considered 'compensatory' for the loss of this final consonant cluster),[43] and it is thus difficult to account for the development of the postposition *kanssa* to the suffix *-kAA* by appealing only to regular rules of phonetic erosion.

[41] Constituents in compound forms retain their original back or front vowels.
[42] "Such applications of vowel harmony across word boundaries generally apply to words such as pronouns, adverbs, conjunctions, and so on, that would ordinarily not receive primary stress in fast speech." Skousen (1975:53f.); Rapola (1966:398).
[43] cf. the development of the Estonian comitative: **kansa* > *-kāsa* > *-kās* > *-ka* > *-ga*. Oinas (1961:16-35).

The introduction of the excessive is not easily described in terms simply of regular phonological changes, either. In the first place, unlike the comitative, the excessive form does not come from a postposition. Instead, the original theory is that the excessive has developed from the case stacking (§3.1.3) of the essive and the partitive exponents *-nA* + *-tA*,[44] with the subsequent loss of the short *-A-* between *-n-* and *-t-*. Such a development is phonologically unusual in Finnish, however, and later scholars have instead preferred to analyse the development of this form as one of analogy (discussed in §3.1.3).

The ongoing phonological developments found in the modern Finnish dialects explain some of the changes that are found in the case system. However, phonetic erosion does not fully explain the introduction of two new case-forms, nor account for the ongoing loss of the abessive, comitative, prolative, and instructive forms. The apparent stability of the 'core' of the Finnish case system may be partly explained by the particular types of phonetic erosion are not leading to any emergent instances of syncretism. However, as will be demonstrated in reference to Veps in §3.3.2, the loss of phonological distinctions does not necessarily always lead to the total loss of morphologically distinct case values.

3.1.3 THE DEVELOPMENT OF THE FINNISH CASE SYSTEM - ANALOGY

Unlike the new comitative form in Finnish, the excessive form *-ntA*, restricted to certain lexemes and to certain dialects, cannot easily be described as having developed phonologically from postpositional material. Instead, we might suggest that it is the result of the analogical extension of a sub-word pattern exhibited by certain local case-forms.[45] In the Finnic languages, the traditional view is that the local case-forms developed by means of *suffixaufnahme* or 'case-stacking', where an element **s* or **l* 'combined' with different direction/place markers:

[44] Lönnrot (1841:35), Ahlqvist (1863:109f.).
[45] Setälä (1891:125f.), Kettunen (1929:26), Ojajärvi (1950:8), and Itkonen (1966:274).

		Local marker	Direction/place marker	Combination	Modern Finnish forms
	Stasis		**nA*	**snA*	*-ssA*
Internal	**Source**	**s*	**tA*	**stA*	*-stA*
	Goal		**n*	**s(V)n*	*-sVVn/-hVn/-Vn*
	Stasis		**nA*	**lnA*	*-llA*
External	**Source**	**l*	**tA*	**ltA*	*-ltA*
	Goal		**n*	**l(V)n*	*-lle(')*

Table 3.5: The traditional view of the development of the Finnish local case-forms[46]

The structure of these case-forms remains transparent in modern Finnish. In slow speech in particular, the source forms are consistently characterized by a dental *-t-* followed by a short final vowel *-A*; the stasis forms by a long consonant followed by a short final vowel *-A*; all the external local case-forms by a form *-l-*; and two (in some inflexion classes three) of the internal local case-forms by a form *-s-*. In the excessive, we find the analogical extension of the synchronically-transparent implicational patterns that exist between the source and stasis case-forms, leading to the production of a novel case exponent *-ntA* which provides a source 'equivalent' for the stasis essive form *-nA*:

adessive : ablative	**inessive : elative**
llA : ltA	*ssA : stA*
nA : ?	*nA : ?*
? + ntA	*? = ntA*

Table 3.6: The analogical development of the excessive case-form

The semantic parallels between the source and stasis forms in the internal and external local sub-systems will have further facilitated the introduction of the excessive case value, although the essive only expresses 'stasis' functions in a handful of forms, e.g., *koto-na* 'at home', *vuon-na* 'in the year', *taka-na* 'behind' (inflected postposition). This may partly explain the lexical restriction of the excessive, as, where the essive form expresses a 'state', it does not participate in the same semantic relationship of 'stasis vs. source' as the inessive and

[46] Hakulinen (1979), Abondolo (2006b:20f.), and Collinder (1960). More recent scholarship suggests that the local case-forms have developed from the suffixation of postpositions, e.g., Aikio and Ylikoski (2007, 2016) and Ylikoski (2016).

elative, or the adessive and ablative. In postpositions such as {TAKA-} 'behind', the 'source' equivalent of the form *taka-na* is *taka-a* 'from behind',[47] which is syntagmatically more opaque than the excessive form *taka-nta* 'from behind', a form which includes the 'internal' marker *-n-* found in the essive and the separative element *-tA* found in the other source forms *-ltA* (ablative) and *-stA* (elative). The transparent sub-word structure of the novel excessive marker may have aided in its development, but the restricted nature of its 'stasis' semantics is likely to have prevented its generalization through the language.

Another development that might be attributed to analogy is the shifting of lexemes between declension classes. For instance, *s*-stem nominals (i.e., those nominals with a nominative singular form ending in *-s*) follow one of three different inflexional patterns. Data gathered for L1 and L2 speakers of Finnish indicate that many words which historically belong to one inflexion class are sometimes deduced 'incorrectly', and assigned to an unetymological class. Nominals historically declining like *vieras* are in some instances starting to follow the pattern of an *ajatus*-type nominal;[48] new loanwords are also 'attracted to' the *ajatus* inflexion class. The change of *s* > *ks* is not the result of regular phonological developments, and must instead be analysed as analogical:

	-Vs → -Vkse- **{AJATUS} 'thought'**	***-Vs → -VV-*** **{VIERAS} 'guest; stranger'**	***-(U)Us → -(U)Ude-*** **{RAKKAUS} 'love'**
Nominative singular	*ajatus*	*vieras*	*rakkaus*
Genitive singular	*ajatuksen*	*vieraan*	*rakkauden*
Partitive singular	*ajatusta*	*vierasta*	*raukkautta*
Essive singular	*ajatuksena*	*vieraana*	*rakkautena*
Inessive plural	*ajatuksissa*	*vieraina*	*rakkauksissa*

Table 3.7: *s*-stem nominals in Finnish

[47] The final *-A* is etymologically the same as the partitive *-(t)A* and originates in a 'separative' marker reconstructed for proto-Uralic. Abondolo (2006b:18-20).
[48] Hakanen (2006).

Unlike in Latin (§2.1.3), this type of analogical development in Finnish is not leading to the loss or propagation of the case system, as the same values are expressed albeit with a different lexical stem shape; the majority of the case suffixes remain the same across declension.

Intriguingly, some forms do not appear to change their inflexion class, but rather exhibit morphological alternations in just one or two paradigm cells. For example, trisyllabic nominals ending in *-eA*, such as the adjective {TÄRKEÄ} 'important', can have two partitive singular forms: *tärkeä-ä* and *tärkeä-tä*.[49] The etymological partitive (singular) form is *-tA*, although after unstressed syllables intervocalic */t/* has generally undergone a softening process following the secondary stressed syllable, first becoming /ð/ and then disappearing.[50] Both *tärkeää* and *tärkeätä* are therefore etymological, though the latter may be considered more archaic. Strict phonological developments are likely to have led to forms in *-A* generalizing. However, the sub-word unit *-tA* is found elsewhere (usually following stems ending in consonants where it did not weaken to /ð/) in the language, and can be easily associated with partitive functions in both the singular and plural. This direct form~function mapping may have facilitated the retention and the spread of the partitive singular exponent *-tA* as an alternative for the form *-A* in nominals ending in *-eA*. This development cannot be considered part of an analogical shifting of lexemes between inflexion classes as it is not accompanied by any other changes (such as, for instance, alterations in the shape of the lexical stem as with the *s*-stem nominals described above). Instead, it involves the analogical generalization of a sub-word unit which can be consistently associated with 'partitive' functions.

[49] Karlsson (2008:124).
[50] Rapola (1966:139).

In these instances, analogy offers some complementary explanations for some of the earlier and ongoing changes found in the Finnish case system.[51] It might even be suggested that certain analogical changes have helped Finnish to maintain its case system. For instance, the phonetic erosion of the *-t-* element in the partitive form *-tA* is to a limited extent reversed by the analogical extension of the form *-tA* from other nominal types. This sort of 'reinforcement' might support the maintenance of case distinctions that otherwise would have disappeared as the result of regular sound change.

3.1.4 THE DEVELOPMENT OF THE FINNISH CASE SYSTEM - SYNTAX

The sort of time-depth presented for the Italic languages in chapter two is not available in reference to any of the Finnic varieties under consideration here. It is therefore not possible to offer similar observations about changes to word order or about shifts towards (non-)configurationality as were offered in §2.1.4, since syntactic developments of this sort may only be detected over long periods of time. Nevertheless, certain syntactic considerations are important for this study, and thus, as far as they may be addressed here, I will offer an account of some of the developments which may have affected the the Finnish case system.

In the first place, a major syntactic difference between the Finnic and Italic languages is the preference for postpositions over prepositions in the former.[52] In the southern Finnish dialects, for instance, the development of the new comitative form is facilitated by the availability of a postposition which occurs in the same position, relative to the noun, as an inflexional case marker. Since in many languages new morphological case-forms develop from postpositions,[53] their presence is frequently a requirement for the grammaticalization of

[51] The importance of transparent sub-word structure in the development of new case suffixes will be discussed further in §6.4.
[52] Intriguingly, all Finnic languages have both postpositions and prepositions, although in all varieties there are far more of the fewer than the latter (prepositions constitute no more than 25% of total adpositions in any Finnic variety). Grünthal (2003:57-59); (2005:31).
[53] Kulikov (2011:440).

novel inflexional suffixes. In the Italic languages, since there were very few postpositions, there were far fewer opportunities for these to develop into new case-forms.[54]

However, the presence of postpositions in a language is not sufficient to determine that a variety will increase its case system. Finnish, for instance, exhibits only one new form from a postposition, and the traditional view is that its other case-forms did not originally develop from postpositions, either (discussed in §3.1.3 in reference to the internal and external local case-forms). It is not clear, either, why Finnish has only developed one new case-form from a postposition, whilst Veps has developed (or is in the process of developing) at least eight new forms from postpositional material (§3.3.4).

There is a similarity across the Finnic and Italic varieties, however, in that they all exhibit both analytic and synthetic means of distinguishing certain functions. There are some instances in which adpositions can encode very similar functions to certain case-forms, but, in a similar way to Latin (§2.1.4), they tend to exhibit slight differences in meaning or to be distributed in different ways. For instance, in Finnish the beneficiary is often expressed by the allative case in *-lle*:

3.2.a *Maailma on tehty me-i-lle*

world.NOM.SG. is made we-PL.-ALL.

'The world was made for (all of) us'

But a similar meaning can be encoded via the use of the postposition *varten* 'for' which (unlike most postpositions) governs the partitive:

[54] One counterexample is found in the Latin pronominal paradigm, where a new comitative form in *-cum* has developed from the suffixation of an adposition of the same shape, meaning 'with'.

3.2.b *Maailma on tehty me-i-tä varten*

world.NOM.SG. is make we-PL.-PART. for

‘The world was made for (the e.g. both of) us’

Although on the surface the allative case and the postposition *varten* seem to govern the same meanings, there is actually a nuanced distinction, where in 3.2.b *varten* gives in the impression that the world was made specifically for a certain number of people (where the pronoun *me* ‘we’ might refer specifically to two or three people, for instance); in the first example, *maailma* ‘the world’ was made for a greater, less specific number of people.[55]

There are also many spatial adpositions in Finnish which seem to encode similar meanings as some of the local cases. Once again, however, there are nuances in the use of synthetic and analytic case-markers:

3.3.a *Istu-n sohva-lla*

sit-1SG sofa-ADE.SG.

’I sit on the sofa’

3.3.b *Tavara-t ovat sohva-n pää-llä*

thing-NOM.PL. are sofa-GEN.SG. on.top.of-ADE.SG.

‘The things are on the sofa’

In these examples, the alternation between the adessive case-form and the use of an inflected form of the postposition {PÄÄ} is constrained by animacy: when someone sits or moves onto the sofa of their own volition, the adessive case is used; when something is

[55] The form in the allative is perfectly grammatical, but the combination of the partitive and postposition *varten* is preferred. Professor Ulla-Maija Forsberg (p.c.).

transferred onto the sofa, the postposition {PÄÄ} is used. In Finnish (as in Latin), adpositions are used to express slightly different meanings than those discriminated by case suffixes. It is therefore difficult to suggest that the restricted use of the abessive, prolative, and comitative case suffixes is the direct result of the presence of postpositional forms encoding the 'same' functions,[56] just as it was problematic to argue that the loss of the Latin case system was caused by the presence of prepositions which discriminated similar functions to (certain) case-forms.

3.1.5 THE DEVELOPMENT OF THE FINNISH CASE SYSTEM - CONTACT CONSIDERATIONS

The sociolinguistic situation in which Finnish has been spoken through time is less clear than that in which Latin has developed. Whilst many historians have noted many of the peoples and places in which Latin was spread and the languages with which they were in contact (even if exact details about the effects these languages had on Latin and its case system are unclear), there is much less information of this sort on the history of Finnish. In order to determine which languages Finnish has been in contact with, we must generally rely on clues found in the language itself rather than on the observations of historians and other written evidence (as is possible to a greater extent with the Italic languages).[57]

Loanwords provide insight into the earliest contact of the Finnic languages. Nouns such as *härkä* 'ox', *vuohi* 'goat', *paimen* 'shepherd', *sisar* 'sister', and *morsian* 'bride' were borrowed during an earlier stage of the Baltic languages (cf. modern Lithuanian *žirgas* 'ox', *ožys* 'goat', *piemuo* 'shepherd', *sesuo* 'sister', and *marti* 'daughter-in-law');[58] nouns such as *kuningas* 'king' and *ranta* 'shore' demonstrate the contact with Germanic (cf. proto-

[56] The instructive is exhibiting a slightly different development, whereby the functions once encoded by the suffix *-(i)n* are now frequently expressed by the adessive, and not an adposition. Similarly, however, from a functional perspective this development is difficult to explain, as there are nuanced differences between the use of the instructive and adessive case-forms.
[57] Referring to the language itself in order to determine what effects contact has had on Finnish is in very real danger of producing a circular argument.
[58] Larsson (2001:239).

Germanic **kuningaz* ‘king’ and modern German *strand* ‘beach’). Later, Finnish came into contact with Russian; the result is extensive lexical borrowing into Finnish (e.g., *ikkuna* ‘window’ from окно (okno) ‘window’ and *raja* ‘border, edge’ from край (kraj) ‘edge’).[59] More recently, during the time particularly when Finland was under Swedish rule, many loanwords came in from the language of the conquerors, such as *ankka* ‘duck’ from *anka* and *kartta* ‘map’ from *karta*.[60] It is therefore towards the Baltic, Germanic, and Slavonic languages that we may look for explanations for certain developments found in the Finnish case system (in particular: why this variety has maintained a large number of morphological case distinctions through time).

As has already been briefly discussed in §2.1.5, according to work on sociolinguistic typology conducted by Thomason and Kaufman (1988) and Trudgill (2011), in contact situations there tend to be patterns of interference, which are often correlated with different attitudes towards and uses of the languages in contact with one another. For instance, when Swedish was transported to Finland as the language of the occupiers, those who natively spoke Finnish tended to learn Swedish. In this way, there was for a long time an uneven direction of language acquisition in the contact situation. Whilst many words were borrowed from Swedish into Finnish, therefore, it is hypothesized that Finnish speakers may have transferred linguistic elements into their L2.[61] According to the typological hypotheses of Trudgill (2011), similar morphological and syntactic interference from Swedish into Finnish should not have occurred to the same extent as the same sort of interference in Swedish, since there were fewer Swedes learning Finnish as a second language. It is therefore difficult to talk about specific changes which occurred in the Finnish case system which may be safely

[59] Häkkinen (2004:250 and 1017 respectively).
[60] Häkkinen (2004:54 and 368 respectively).
[61] Thomason (2001:chapter four).

attributed to the influence of Swedish, even though this contact situation occurred relatively recently.

It is even more difficult to delimit for certain the earlier effects of contact with the Baltic and Germanic languages, as reliable data are much more limited than they are for the more recent effects of contact. Nevertheless, judging from the abundance of loanwords from Baltic and Germanic in Finnish, it may be surmised, according to the typology of contact-induced changes, that morphological and syntactic interference was more likely to have occurred in the contact varieties than in Finnish. Without further information regarding the nature of the contact situation, not a great deal more detail can be given about the specific influence of these languages on the Finnish case system. Nevertheless, there are a couple of developments found in Finnish which are generally analysed to be the result of contact.

In the first place, further to extensive loaning of lexical items from Baltic, Slavonic, and Germanic languages, it has been suggested that prepositions have been introduced into the Finnic languages through contact with varieties which are also characterized by prepositional marking.[62] Grünthal (2005:37) explains that there are two main weaknesses with the assumption that prepositions were introduced as a functional category to the Finnic languages by contact particularly with neighbouring Baltic and Germanic varieties. The first is that most Finnic languages lack borrowed prepositions completely (and even those that exhibit them have very few).[63] The second is that the morphosyntactic structure of the prepositional phrase is completely different from that of the postpositional phrase: the 'proto-typical' structure of the latter is [[NOUN-gen.] + post.] whilst that of the former is [prep. + [NOUN-part.]]. The prepositional phrasal structure of the languages with which the Finnic languages have most closely been in contact (e.g., German, Latvian, Russian) does not involve a 'partitive' case-

[62] The development of prepositions in Finnic tends to be regarded as an innovation. Grünthal (2005).
[63] There are a maximum of four in Veps (Kettunen 1943:538, 544f.) and two in Livonian (Grünthal 2003:177-200).

form (since these languages do not have one), so it is not clear how this distinct prepositional structure should have been introduced into Finnish and other Finnic varieties.[64]

After concluding that the introduction of prepositions cannot be explained as the result of language contact alone, Grünthal suggests that the development of adpositions which originally occurred after a noun or noun phrase (or which could be bipositional) into prepositions followed the word order change S(ubject)O(bject)V(erb) to S(ubject)V(erb)O(bject).[65] This is another syntactic change hypothesized to have occurred in the Finnic varieties as a result of language contact, since many of the Baltic, Slavonic, and Germanic languages with which Finnic was in contact were (predominantly) SVO;[66] the basic word order of proto-Uralic is generally reconstructed as SOV.[67] Furthermore, the Finnic languages exhibit a number of morphological and syntactic properties which are typically found in languages with SOV word order, including postpositions, very few prefixes, and a suffixing preference.[68] The Finnic languages, however, exhibit basic SVO word order.[69] If the basic word order of the Finnic languages has undergone a change from SOV to SVO, it may help explain the development and increase in the number of prepositions. However, since the reconstruction of SOV word order relies on abductive reasoning, rather than on earlier attestations of related varieties, it is difficult to determine: a) whether Finnic *did* undergo a

[64] Moreover, even if contact is the sole reason for the increased use of prepositions, other Uralic languages in contact particularly with Indo-European languages (which are largely prepositional) would surely have exhibited a similar shift. Since this is not what we find, the least that must be suggested is that there was some aspect of the Finnic languages which made them more susceptible to the influence of the neighbouring Germanic, Baltic, and Slavonic varieties and their preference for prepositions over postpositions. Ehala (1994) discusses this hypothesis in reference to Estonian.
[65] 2005:39.
[66] Many of the arguments that appeal to contact as the source of a change SOV to SVO are overly simplistic. Although a 'basic' word order may sometimes be identified, there are frequent differences between, for instance, main and subordinate clauses (e.g., German, Estonian).
[67] Although "since subject and object noun phrases are regularly omitted in Uralic sentences it is probably more helpful to speak of '(T)FV', i.e. (optional) Topic(s)-Focus-Verb order." Abondolo (2006b:33).
[68] Greenberg (1963).
[69] Finnish word order is more accurately described as being determined by the order of the 'topic' and 'focus'. Vilkuna (2003).

change SOV to SVO, b) whether this development was the result of language contact, and c) whether it led to the introduction and increased use of prepositions.

Without more detailed information about the nature of the sociolinguistic situations in which Finnish has been used and learned, it is very difficult to offer any further details about the changes that might have occurred in the language as the result of contact with other linguistic varieties, particularly those which have affected the case system. Even less, it appears, may be said with any confidence about the sociolinguistic effects on Finnish than was offered in reference to the Italic varieties in the preceding chapter.

3.2.0 INTRODUCTION TO THE LIVONIAN LANGUAGE

As a living language,[70] Livonian was spoken in the Courland peninsula and along the north coast of Latvia.[71] Two dialectal variants of Livonian are usually distinguished: Courland (Latvian *Kurzeme*) and Salaca (Latvian *Vidzeme*). The former is usually split further into three 'sub-dialects': east (spoken in the villages of Ūžkilā, Sīkrõg, Irē, Kuoštrõg, Pitrõg, Sānag, Vaid, Kūulka, and Mustānum), central (Īra) and west (Lūž and Pizā).[72] Courland Livonian, the best documented Livonian dialect and the basis of the written language,[73] will be focus of this study, although certain differences found in Salaca Livonian will be referred to where relevant.

In reference to Livonian, *ḑ*, *ţ*, *ļ*, *ŗ*, *ņ*, *ŗ*, *š*, and *ž* are used for palatalized consonants; *õ* for the high central vowel; *ȯ* for the mid-high back vowel, and *ǭ* for the (long) mid-central vowel for which there is no short equivalent.[74] I will use a macron to distinguish long vowels from short, whilst long consonants will be written as geminate. Finally, an apostrophe is used to

[70] The last native speaker of Livonian died in 2009. Kittilä and Ylikoski (2011:48).
[71] Grünthal (2015b:16f.).
[72] "The main difference is between East and West Livonian, while Central Livonian forms a transition area." Tuisk (2016:122).
[73] Pajusalu (2014:149f.).
[74] The precise phonetic value of particularly *õ* varies depending on the dialect. Pajusalu (2014:154-161).

indicate a ‘broken tone’ “which is rising-falling or predominantly falling and articulated with laryngealization (stød or creaky voice)”.[75] The plain tone is not marked, and occurs on primary stressed syllables (i.e., the first syllable) where the broken tone does not.

3.2.1 THE LIVONIAN CASE SYSTEM

Livonian has the fewest morphological case-forms of any Finnic language, with grammars usually listing between eight and twelve values:[76]

	{LÄPŠ} ‘child’		**{KALĀ} ‘fish’**	
	Singular	**Plural**	**Singular**	**Plural**
Nominative	*läpš*	*lapst*	*kalā*	*kalād*
Genitive	*laps*	*lapst*	*kalā*	*kalād*
Partitive	*lapsta*	*läpši*	*ka’llõ*	*ka’ļḍi*
Dative	*lapsõn*	*lapstõn*	*kalān*	*kalādõn*
Essive	*lapsõn*	*lapstõn*	*kalān*	*kalādõn*
Translative	*lapsõks*	*lapstkõks*	*kalāks*	*kalādõks*
Elative	*lapsõst*	*läpšīst*	*kalāst*	*ka’ļšti*
Inessive	*lapsõs*	*läpšīs*	*kalās*	*ka’ļši*
Illative	*lapstõ*	*läpšīz*	*ka’llõ*	*ka’ļži*
Comitative	*lapsõks*	*lapstkõks*	*kalāks*	*kalādõks*

Table 3.8: Livonian {LÄPŠ} and {KALĀ} fully declined

The functions of the essive case value are frequently encoded by a form the same as the dative;[77] an instructive case is occasionally counted, but it has completely fossilized.

The first striking difference between the Finnish and Livonian case systems is the fact that, in the latter, not all non-nominative case-forms are distinguished by overt affixal marking: the genitive singular form, where it is distinct from that of the nominative singular, is only distinguished by means of morphophonological differences in the lexical stem. The Livonian case system is also characterized by more instances of syncretism than other Finnic varieties. In all nominals, the nominative plural and genitive plural functions are encoded by a single

[75] Viitso (2007:47).
[76] Viitso (2016:146) counts twelve; Grünthal (2003) lists eight where the translative-comitative is considered a single form encoding ‘translative’ and ‘comitative’ meanings.
[77] Viitso (2016) identifies a small number of lexemes in which the essive and dative forms are distinct.

form; in all but a handful of nominals in Courland Livonian, the comitative and translative functions are encoded by the same form (distinct across number);[78] and, in many nominals, the partitive singular and illative singular functions are not distinguished phonologically.

In Courland Livonian, the external local cases (ablative, adessive, and allative) are no longer productive, but are "now restricted to farmstead names, certain village names, place-names ending in *mǭ*, to a few nouns (and their adjectival attributes) [...] and to some adverbs and postpositions";[79] there is no longer any morphological distinction between the adessive and allative even in these lexemes.[80] The functions once encoded by these forms are now largely expressed analytically or by the new dative form. Courland Livonian is also the only Finnic variety to have a dative case, the origins of which are unclear,[81] whilst in Salaca Livonian, there is no morphological dative case-form. Instead, 'beneficiary' and 'recipient' meanings are encoded by a productive adessive-allative form; comitative and translative values are consistently encoded by distinct forms.[82]

The main uses and functions of the Livonian cases are as follows:

- nominative: the form of the telic subject; unlike in Finnish, this is always the form of the predicative adjective, and never that of the object;[83]
- genitive: used both adnominally (to express possession) and adverbally (as the telic direct object);
- partitive: used as the form of the atelic, indefinite subject as well as the atelic object;

[78] Although compare the following forms: *naizõks* 'for the wife' vs. *naizkõks* 'with the wife', *sūrõks pi'ņņõks* '[grows] into a big dog' vs. *sūr piņkõks* 'with a big dog'. Viitso (2016:147). Here, I will generally refer to a single 'translative-comitative' form in Courland Livonian.
[79] Viitso (2016:147).
[80] Ernštreits and Kļava (2014:80).
[81] Inaba (2015) offers a comprehensive account of the Courland Livonian dative and its possible historical roots.
[82] Pajusalu (2014:162f.).
[83] Viitso (2012:21).

- dative: used to mark the ‘recipient’, ‘beneficiary’, and, sometimes, ‘source’;
- translative-comitative: encodes ‘sociative’, ‘translative’, and ‘instrumental’ meanings;
- internal local cases: encode similar functions as the internal local case-forms found in Finnish (illative ‘into’; inessive ‘in’; elative ‘out of’).[84]

Most Livonian nominals exhibit allomorphy in the stem, though the extent to which the stems differ from one another phonologically varies across the system. We find some examples of invariant stems (e.g., {KĪNDÕR} ‘elbow’), stems that differ solely in the palatalization of their final consonant before *-i* (e.g., {PĒGAL} ‘thumb’), and many lexemes that undergo a ‘grade alternation’, which

> “concerns words having both a short nuclear vowel and a heavy coda in the first syllable of strong-grade forms. A heavy coda is produced with the broken tone or it contains a phonetically half-long or full-long vowel or consonant in syllables with the plain tone. In weak-grade forms coda is either absent or light. Weak-grade forms have a long vowel in the second syllable if the first syllable is short or in the first syllable if this syllable is long.”[85]

Gradation can therefore involve alternations in: the length of a stem consonant (*ka<u>l</u>ād ~ ka’<u>ll</u>õ*); the length of the stem vowel in the first syllable (*l<u>ī</u>lõd ~ l<u>i</u>llõ*); the length of the stem-final vowel and that of the case(/number) ending (*aig<u>ā</u>d ~ a’ig<u>õ</u>*); the quality of a vowel in the first syllable (*<u>ō</u>kõd ~ <u>ou</u>kõ*), and in tone (*aigād ~ a’igõ*). Many lexemes exhibit more than one of these alternations:

[84] Viitso (2016).
[85] Viitso (2007:45).

Nom. sing.	Nom./gen. pl. - Weak stem	Part. sing. - Strong stem	Gloss
kalā	*kalā-d*	*ka'll-õ*	fish
aigā	*aigā-d*	*a'ig-õ*	edge; shore
lil	*līlõ-d*	*lill-õ*	stem of an umbellifer
ouk	*ōkõ-d*	*ouk-õ*	hole; pit

Table 3.9: Comparing weak and strong stems in selected Livonian nominals[86]

Quantitative grade alternation (that affecting the length of vowels and consonants) is the most common type found in Courland Livonian. There are fewer instances of qualitative grade alternation in this dialect, although in Salaca Livonian it is more common.[87] Importantly, on consideration of the extensive discussion in Viitso (2007), it does not appear as though it can be predicted with certainty whether a lexeme exhibits grade alternation, or what the different stem shapes of a nominal are, from knowledge of a single form alone. The phonological differences exhibited by weak and strong stems seem to have morphologized in the modern language.

Formal variation in the case suffix across both number and declension is also a great deal more extensive in Livonian than in Finnish:

	Singular	Plural
Nominative	*Ø*	*-d, -õd, -t*
Genitive	*Ø*	*-d, -õd, -t*
Partitive	*-tā, -da, -ta, -ţa, -dõ, -tõ, -t, -õ, -i*	*-di, -ti, -ḍi, -ţi, -i, -īdi, -idi*
Dative	*-n, -õn*	*-ddõn, -dõn, -õdõn, -tõn*
Elative	*-stõ, -st, -õst*	*-šti, -īst(i), -ist(i)*
Inessive	*-sõ, -s, -õs, -š(õ)*	*-ši, -īs(i), -is(i)*
Illative	*-zõ, -õ(z)*	*-ži, -īž, -iž, -ž, -īz, -iz*
Translative/comitative	*-kõks, -ks, -õks*	*-dkõks, -tkõks, -dõks, -õdõks, -tõks*

Table 3.10: Distinct Livonian case exponents[88]

[86] Table adapted from Viitso (2007:46).
[87] Pajusalu (2012:207).
[88] Table adapted from Viitso (2012:22). Pronominal suffixes included in the original table have been omitted here.

There is no productive vowel harmony in Livonian, although there are (rare) instances in which certain suffixal desinences appear to be distributed according to phonological considerations. For instance, the dative singular has two exponents: *-n* and *-õn*. Regularly through the language, *-n* is found with lexical stems ending in a vowel, whilst *-õn* occurs with lexical stems ending in a consonant. Furthermore, the internal local case-forms in both the singular and the plural exhibit a 'mobile' *õ* which can occur before or after the sibilant *s/z* (or palatalized equivalent). Generally, where a case value is characterized by the forms *-õC(C)* and *-C(C)õ*, the former occurs with a lexical stem ending in a consonant whilst the latter is found with a lexical stem ending in a vowel. However, the form *-C(C)õ* is also found with certain consonantal stems,[89] which means that there is no strictly phonological environment which consistently licenses the introduction of either the *-õC(C)* or *-C(C)õ* suffixal shape.

The exact number of declension classes in Livonian is not generally agreed upon: the Livonian dictionary lists thirteen 'major' classes and 242 sub-classes of nominals (nouns, adjectives, numerals, demonstratives and pronouns);[90] Viitso (2012) identifies eight major classes and 85 'sub-classes'. As in Finnish, Livonian declensions are generally differentiated according to distinct stem shapes and their distribution within the inflexional paradigm; sub-classes are those which follow the 'template' of the declension class, but which differ slightly in terms of the precise morphophonological differences are exhibited by stem allomorphs. For instance, the lexemes *kalā*, *tuba*, and *aigā* all belong to the same declension class, and yet each exhibits slightly alternations in the shape of the lexical stem:

[89] Viitso (2012:23).
[90] Viitso and Ernštreits (2012).

	{KALĀ} 'fish		{TUBA} 'room; house'		{AIGĀ} 'edge'	
	Singular	**Plural**	**Singular**	**Plural**	**Singular**	**Plural**
Nominative	*kalā*	*kalād*	*tubā*	*tubād*	*aigā*	*aigād*
Genitive	*kalā*	*kalād*	*tubā*	*tubād*	*aigā*	*aigād*
Partitive	*ka'llõ*	*ka'ļḍi*	*tu'bbõ*	*tu'dḍi*	*ai'gõ*	*a'igḍi*
Dative	*kalān*	*kalādõn*	*tubān*	*tubādõn*	*aigān*	*aigādõn*
Elative	*kalāst*	*ka'ļšti*	*tubāst*	*tu'bšti*	*aigāst*	*a'igšti*
Inessive	*kalās*	*ka'ļši*	*tubān*	*tu'bši*	*aigās*	*a'igši*
Illative	*ka'llõ*	*ka'ļži*	*tu'bbõ*	*tu'bži*	*ai'gõ*	*a'igži*
Translative-comitative	*kalāks*	*kalādõks*	*tubāks*	*tubādõks*	*aigāks*	*aigādõks*

Table 3.11: Livonian nominals {KALĀ}, {TUBA}, and {AIGĀ} fully declined[91]

In *kalā*-type nominals, the stem consonant -*l*- lengthens before the inflexional exponent in the partitive and illative cells in the singular (with attendant alternation in tone); this consonant palatalizes in the partitive, elative, inessive, and illative cells in the plural. In the paradigm of {TUBA}, on the other hand, whilst there is lengthening of the stem-final consonant and the alternation in tone, in the partitive, elative, inessive, and illative cells in the plural the stem-final consonant -*b*- does not palatalize. Finally, in the paradigm of {AIGĀ}, there are no length alternations in the stem-final consonant in the relevant singular forms and no palatalization in the relevant plural forms. Nevertheless, the distribution of these different stem shapes and suffixal exponents is the same for each of these lexemes:

	Singular	**Plural**
Nominative	[STEM1]	[STEM1]-*d*
Genitive	[STEM1]	[STEM1]-*d*
Partitive	[STEM2]-*õ*	[STEM2]-*ḍi*
Dative	[STEM1]-*n*	[STEM1]-*d*-*õn*
Elative	[STEM1]-*st*	[STEM2]-*šti*
Inessive	[STEM1]-*s*	[STEM2]-*ši*
Illative	[STEM2]-*õ*	[STEM2-]-*ži*
Translative-comitative	[STEM1]-*ks*	[STEM1]-*d*-*õks*

Table 3.12: Paradigm template for *kalā*-, *tuba*-, and *aigā*-type nominals

[91] Strong stems shaded.

All lexemes which belong to this inflexion class follow these patterns, and yet may differ in the ways in which their stem shapes alternate.[92]

3.2.2 THE DEVELOPMENT OF THE LIVONIAN CASE SYSTEM - PHONOLOGY

Livonian is often considered the most phonologically 'eroded' of the Finnic languages. It has lost a number of morphological case-forms that are found in related varieties, and it is often suggested that this is at least in part the result of extensive phonological developments. The total loss of the genitive singular marker *-n*, for instance, has led to the syncretism of the nominative singular and genitive singular in many Livonian nominals, although the instances of morphophonological alternations in the lexical stem mean that the genitive singular form is still occasionally distinct from the nominative singular form.

In the plural, the nominative and genitive functions are always expressed by the same form. Some have argued that this development is at least partially the result of regular phonological changes. For instance, Kettunen suggests that the genitive plural first lost the 'connecting vowel' between the *-t-* and *-n-* of the suffix **-ten* as the result of regular phonetic erosion; final **-n* would then have been lost following the same development which affected the genitive singular form. However, he argues that further changes must be explained by making reference to the analogical 'interference' of the nominative plural form.[93] For instance, for the lexeme {LÄPŠ} 'child', the form ***laśt* (compare Finnish genitive plural *lasten*) would be expected form of the genitive plural. Instead, we find the form *lapśt*, a form identical to that of the nominative plural.

Courland Livonian is the only Finnic language in which 'comitative' and 'translative' functions are expressed by the same case-marker. The traditional analysis is that two

[92] An easily-accessible and up-to-date list of Livonian inflexion classes can be found on the website for the Livonian dictionary: http://www.murre.ut.ee/liivi/noomenityybid.html#17.
[93] 1947:58.

originally distinct forms which 'merged' phonologically. The translative suffix *-*ksi* lost its final *-*i* via regular sound changes (cf. the ongoing loss of final -*i* in modern Finnish), whilst the comitative -*ks* originates from a postposition **kansa*: **kansa* > *-*kās* > -*kas* >-*kəs* > -*ks*.[94] However, Grünthal notes that some of the posited phonological changes are difficult to motivate in light of other evidence:

> "Although Livonian is the most eroded Finnic language in many ways and the attrition of word-final elements has been intensive, the assumption that the original stressed first-syllable *a* would have been dropped is simply very unlikely to be true. First, in nouns such as *jālga* 'foot' with a long first syllable, the word-final -*a* has been preserved. Second, the change from a free word to a bound morpheme does not improve the hypothesis, because the assumed intermediate form -*kas* is preserved in other contexts. This is identical with the historical ending of a certain noun type, cf. Livonian *kāngaz* 'cloth, stuff', *rikkaz* 'rich'."[95]

A purely phonological account of the development of the comitative suffix is unsatisfactory, and instead there were likely to be other factors at work, which will be discussed in more detail in §3.2.5.

Courland Livonian has also lost the Finnic external local case series almost entirely; etymological external local markers are restricted to only a few nouns in the modern language (e.g., {KE'Ž} 'hand', {SUODĀ} 'war', {MQ̄} 'land, country').[96] In other Finnic varieties, such as Votic and Karelian, we find occasional reductions in the number of morphologically distinct external local case-forms, but nowhere else do we find the total loss of the ablative, adessive, *and* allative case values from a Finnic language. That fact that certain distinctions are still maintained by a handful of lexemes may indicate that phonetic erosion alone cannot account for the loss of productive external local case suffixes.

[94] Kettunen (1947:60), Kahr (1976:117), Oinas (1961:14).
[95] 2003:181.
[96] Viitso (2016:147). Even in these lexemes, where are no allative, adessive, or ablative forms in the plural. Salaca Livonian maintains a productive adessive-allative form.

Finally, and perhaps most intriguingly, Courland Livonian is characterized by a dative form. Since no other Finnic variety (including the Salaca Livonian dialect) exhibits a morphologically separate dative form, this is invariably analysed as an endogenous innovation. However, the exact history of this form is not known precisely and the textual evidence does not extend back far enough to offer us a glance at its earliest roots. There are nevertheless a number of suggestions for how it may have developed. Some believe that the dative form in *-n* has its origins in the essive form **-nA*; Alvre (1974) suggests that it developed from the lative **-n* ~ **-ń*; yet others hypothesize that it originated from the proto-Finnic genitive **-n*:[97]

> "As final *-n* is generally lost in Livonian, the Finnic marker of the genitive *-n* was preserved only in sandhi-position before vowels as in *m^{i}e'r-n-ajgās* 'on the beach of the sea'. The split of the Finnic genitive in dative (*-n*) and genitive (*-Ø*) in Livonian thus started before postpositions beginning with a vowel [...]. The Livonian dative was thus originally only a (more distinct) variant of the genitive."[98]

Regardless of which of these hypotheses is correct (if, indeed, any of them are), what remains constant is the supposition that this form developed from pre-existing (morpho-) phonological material. In this way, alongside the phonological developments that may have occurred, we find a concomitant shift in the functions originally signalled by the form **-nA/*-n/*-ń*. This sort of functional shift must therefore be included alongside phonological considerations when the development of the Courland Livonian dative is considered.

Although phonological erosion clearly played an important role in the development of the Livonian case system, appealing to regular phonological developments cannot account satisfactorily for the syncretism of the nominative and genitive in the plural, for the 'merging' of the translative and comitative cases, for the introduction of a dative case feature value, or for the total loss of external local case-forms from the majority of nouns and adjectives. In the

[97] Wälchli (2000:214f.) and (2001:430-432), Grünthal (2003:100).
[98] Wälchli (2001:430f.).

introduction of the dative value, for instance, reference to a functional shift is also necessary and, in the development of the unetymological syncretism of the nominative plural and genitive plural forms throughout the language, analogy must have played a part alongside phonological developments. Nevertheless, it is notable that the erosive forces Livonian has undergone have had extensive effects on the case system - in particular, the increase in fusional structure (such as frequent allomorphy in the stem, cumulative expression of case and number values, numerous instances of syncretism, and the lack of affixal marking in the genitive singular). The significance of this change will be discussed further in §6.2.

3.2.3 THE DEVELOPMENT OF THE LIVONIAN CASE SYSTEM - ANALOGY

With the paucity of historical data and the lack of native speakers from whom to gather data, the role of analogy in the development of the case system is more difficult to discuss in reference to Livonian than was possible in reference to Finnish. Nevertheless, it is possible to detect the results of analogy in the paradigms of Livonian lexemes which exhibit overabundance in particular cells:

	{ĀIGA} 'time'	**{SĪLMA} 'eye'**	**{KĀ̆NGA} 'shoe'**	**{Ā̆RGA} 'ox'**
Nom./gen.pl.	*āigad*	*sīlmad*	*kā̆ngad*	*ā̆rgad*
Part. pl.	*aigi*	*siļmi*	*keŋgi*	*eŗgi*
In.pl.	*āigis ~ aigši*	*sīlmis ~ siļmši*	*kā̆ngis ~ keŋgši*	*ā̆rgis ~ eŗgši*
El.pl.	*āigist ~ aigšti*	*sīlmist ~ siļmšti*	*kā̆ngist ~ keŋgšti*	*ā̆rgist ~ eŗgšti*
Ill.pl.	*āigiž ~ aigži*	*sīlmiž ~ siļmži*	*kā̆ngiž ~ keŋgži*	*ā̆rgiž ~ eŗgži*

Table 3.13: Alternating forms in selected Livonian nominals

Each of the plural internal local case-forms of these lexemes is overabundant. That is, two forms occupy single cells where only one is expected. Each of these forms is given in the Livonian dictionary (both the hard-copy and online versions), which indicates that each variant must be 'acceptable' to at least some degree. Phonological changes would have

resulted in a single etymological form in the Courland dialect; the introduction of a second is therefore here analysed as an analogical development.

From consideration of the patterns exhibited by other nominals with a nominative singular in *-a/-ā*, it is clear that these forms are each derived on the model of distinct inflexional patterns, encapsulated by the lexemes {AMĀ} 'all, whole, entire' and {IZĀ} 'father':

	{AMĀ} 'all, whole entire'	**{IZĀ} 'father'**
Nominative/genitive plural	*amād*	*izād*
Partitive plural	*ä'mḍi*	*i'žḍi*
Dative plural	*amādõn*	*izādõn*
Inessive plural	*ä'mši*	*izīs*
Elative plural	*ä'mšti*	*izīst*
Illative plural	*ä'mži*	*izīž*
Translative-comitative plural	*amādõks*	*izādõks*

Table 3.14: {AMĀ} and {IZĀ} fully declined in the plural

For the lexemes in table 3.15 below, there are two plural internal local case-forms, one based on the strong stem (on the model of *izā*-type nominals) and one based on the weak stem (on the model of *amā*-type nominals):

Stem type:	**strong**	**weak**	**strong**	**weak**	**strong**	**weak**	**strong**	**weak**
In.pl.	*āigis*	*aigši*	*sīlmis*	*siļmši*	*kā̄ngis*	*keŋgši*	*ā̄rgis*	*eŗgši*
El.pl.	*āigist*	*aigšti*	*sīlmist*	*siļmšti*	*kā̄ngist*	*keŋgšti*	*ā̄rgist*	*eŗgšti*
Ill.pl.	*āigiž*	*aigži*	*sīlmiž*	*siļmži*	*kā̄ngiž*	*keŋgži*	*ā̄rgiž*	*eŗgšti*

Table 3.15: Inflexional patterns in {ĀIGA}, {SĪLMA}, {KANGA}, and {ĀRGA} plural forms

The implicational patterns that exist between the internal local case-forms and other forms in the paradigm appear to have been analogically extended to lexemes that etymologically exhibit different abstract patterns. Knowledge of one plural internal local case-form provides sufficient implicational information to determine the shape of the remaining internal local case-forms in the plural in these lexemes; where one of these is not known, the speaker has to choose between two equally valid sets of implicational patterns and will sometimes choose incorrectly. The result is overabundance.

Although there are likely to be many other examples of analogical developments in the Livonian case system, the nature of the data available, and of the lack of native speakers, restricts the discussion that can be offered in this section. But its brevity reflects the paucity of the evidence, and not necessarily the number of analogical developments to have affected the Livonian case system.

3.2.4 THE DEVELOPMENT OF THE LIVONIAN CASE SYSTEM - SYNTAX

In addition to the developments discussed in reference to Finnish in §3.1.4 (namely, the putative word order change SOV to SVO and the development, and subsequent increase, of prepositions), in Livonian there appears to be an increase in the number of instances in which the correct interpretation of the functions of constituents in the sentence is determined solely by word order (although it is impossible to claim that this is a direct result of the loss of distinct case-forms in certain lexemes):

3.4.a *un siz se izānd izā kīt-iz*

and then this master.NOM./GEN.SG. father.NOM./GEN.SG. say-IMPF.SG.3

se puoga pāl

it son.NOM./GEN.SG. (up)on

'And then the master's father said to that son/boy'[99]

This example

> "illustrates the most extreme consequences of inflectional reduction. The argument structure and the distinction between grammatical relations are based solely on word order in such cases as the one above. Since the merger of the nominative and the genitive is rather widespread, sentences with the same structure as [3.4.a] are quite common."[100]

The intertwined phonological (loss of *-*n*) and syntactic (rise of configurationality) developments described here are not necessarily strictly *causal* - that is, the loss of the

[99] From Grünthal (2003:96).
[100] Grünthal (2003:97).

morphological distinction of the nominative singular and genitive singular forms of most lexemes is unlikely to have directly led to a rise in configurationality. After all, not all lexemes in Livonian have lost these morphological distinctions, so in certain instances word order would not be the only means of determining the functions encoded by a certain form. Nevertheless, it is notable that we find a similar loss of case-forms and a concomitant increased reliance on word order in the discrimination of constituent functions in two unrelated varieties, Livonian and Latin.

Furthermore, Livonian has not only developed prepositions, but has actually borrowed two prepositions from the neighbouring Latvian, with which it has been in long-term and intense contact (discussed in §3.2.5): *bäs(s)/bäz* 'without' and *pa* 'to'.[101] Cross-linguistically, it is rare for languages to borrow adpositions, and such instances are only found in those languages which are in long-term and intensive contact situations - precisely the type of situation in which Livonian was spoken in the last decades of its life. Intriguingly, the introduction of the preposition *pa* may, Grünthal (2003:177-185) argues, have affected the development of the translative-comitative case, as will be outlined in more detail in the following section.

Finally, external local case functions in the modern Livonian language are encoded by adpositions, not inflexional case-markers. The paucity of the historical data makes it impossible to trace the exact developments of the ablative, adessive, and allative desinences. We may suggest that the availability of adpositional marking facilitated the loss of distinction suffixes, although it would beg the question: why have the internal local case-forms not also fallen out of use, given the presence of an adposition *sizāl* 'inside', for example? Without further evidence, we may observe only that, for the external local case values which are

[101] Grünthal (2003:53).

elsewhere in Finnic expressed synthetically, analytic marking is preferred in modern Livonian.

3.2.5 THE DEVELOPMENT OF THE LIVONIAN CASE SYSTEM - CONTACT CONSIDERATIONS

There are major similarities in the sociolinguistic contact situations in which Finnish, Livonian, and Veps have been spoken throughout time. All three of these Finnic varieties have all been in frequently intense contact throughout history, often with languages which exhibit fewer case distinctions than they do and which are characterized by more fusional morphological structure; moreover, Livonian and Veps have tended not to be the language used for official church or state business. Nevertheless, we find different developments through time, where Livonian has lost, Veps has propagated, and Finnish had tended to maintain its inherited morphological case-forms. Consideration of the historical contact situations alone does not, therefore, provide satisfactory explanation for the changes that are found in these Finnic languages.

In the last few decades before its death, particularly during the years of the World Wars, all native speakers of Livonian were necessarily fluent in Latvian. As a result, we find significant influence from Latvian at every level of the Livonian language: the lexicon, phonology, syntax, and morphology. Post 1920 there was a 'national awakening', during which Livonian was taught at certain schools in the Courland peninsula area, though following the Second World War these activities were suppressed and "[a]ll expressions of Livonian culture were proscribed by the Soviets".[102]

In the case system, potential interference from Latvian may be found in the development of the translative-comitative form in Livonian. As was discussed in §3.2.1, the hypothesis that two originally separate forms - the translative (suffix) and the comitative (postposition) - 'fell

[102] Blokland and Hasselblatt (2003:132).

together' as the result of phonological changes is problematic. An alternative view is that the development of the case-form in *-ks* is largely the result of contact with Latvian. Instead of claiming that two originally separate case-forms merged phonologically, Grünthal (2003) argues that (originally translative) *-ks* developed comitative meanings by means of an analogical extension of functions on the model of the Latvian instrumental case:

- An inflected Livonian nominal ending in *-ks* encodes the following functions:
    1. the translative (or 'predicative adverbial');
    2. the instrumental;
    3. the comitative;
    4. in combination with the preposition *pa*, the translative *only*.
- In Livonian there is a preposition *pa* which has been borrowed from Latvian.[103] In Latvian it governs the (accusative-)instrumental case (in the singular) and can encode the meaning 'to, about, like'.[104]
- In Latvian there is a second preposition *par* which, in certain dialects (crucially, those that were spoken around Courland), merged phonologically with *pa.*
- This preposition *par* in Latvian governs the same (accusative-)instrumental case as *pa* but instead encodes 'predicative adverbial' and also 'predicative instrumental' functions.
- Additionally, the bare (accusative-)instrumental case in an earlier stage of Latvian could be used to express a comitative (sociative) meaning.
- As a result, Grünthal concludes that "[t]he morphosyntactic properties of the TRANSL[ative]COM[itative] are not the result of metaphoric extension, but of a contact-induced analogical change":[105]

[103] Kettunen (1938:269).
[104] The Latvian 'instrumental' is often counted in the modern language as a case, though it is marked by a form identical to the accusative in the singular and the dative in the plural. Mathiassen (1997:41;184f.).

Latvian:		Livonian:
[*pa(r)* + [N + (ACC.-)INSTR.)]] = TRANSL.	:	[*pa* + [N + TRANSLCOM.]] = TRANSL.
[N + (ACC.-)INSTR.)] = COM.	:	[N + TRANSLCOM.] = X
	X :	COM.

Figure 3.1: Contact-induced analogical change of Livonian *-ks*[106]

Even if the Livonian translative-comitative form did not develop in this way, from the discussion offered here and in §3.2.2, it is suggested that making reference to only one type of development (whether that is phonological, analogical, or syntactic, for instance) is frequently not sufficient to explain this (or any other) change exhibited by the Livonian case system. A network of factors needs to be considered in order to best account for the loss of the external local case-forms, the development of a dative form, and the unusual syncretism of translative and comitative functions. However, as a result of the paucity of data, the precise details of these developments cannot be traced.

3.3.0 INTRODUCTION TO THE VEPS LANGUAGE

The name 'Veps' traditionally covers three major dialects: northern, central, and southern - in addition to a written variety. The main Veps dialects have traditionally been spoken in and around the Russian villages "from the southwestern shores of lake Onega and spreading southward over the river Svir' between the lakes Ladoga and Belozero."[107] However, following increased migration particularly to St. Petersburg after the Second World War, native speakers of Veps dispersed and the use of the language decreased drastically.

---

[105] 2003:194.
[106] Adapted from Grünthal (2003:194).
[107] Puura et al. (2013:4).

Of the linguistic study conducted on Veps, the most accessible and comprehensive is that of the written language. For this reason, I will focus on this variety here, although I will make frequent reference to the three major dialects. The term 'Veps' will be used throughout as a shorthand for the written language, whilst the dialects will be referred to explicitly where they are discussed.

3.3.1 THE VEPS CASE SYSTEM

Veps exhibits the greatest number of morphological case distinctions of any Finnic language, with between seventeen and twenty-two inflexional cases usually identified.[108] Table 3.16 includes some of the 'younger' case-forms in Veps which have developed from the suffixation of postpositions or postpostional adverbial particles, including: the source case-forms (from the suffixation of a directional particle *päi*); the approximative local case-forms (from the suffixation of an inflected form of the postposition **lō-*); the terminative (from the suffixation of a postposition **sāδen* 'towards'); the prolative (from the suffixation of a postposition *möto* 'across, via'); and the comitative (from the suffixation of an inflected form of the postposition **kerta* 'with'). These forms exhibit behaviour that differs slightly from the 'older' case-forms, and will be described in §3.3.3.

[108] Grünthal (2015a) lists seventeen; Tikka (1992) and Kettunen (1943) omit the instructive and include discussion on an additional two sets of three local case-forms: the superessive and subessive. These are not included here since they have not grammaticalized to the same extent as other 'younger' case-forms (§3.3.3).

	{SUGA} 'comb'		{KÄZI} 'hand'	
	Singular	**Plural**	**Singular**	**Plural**
Nominative	*suga*	*suga-d*	*käzi*	*käde-d*
Genitive	*suga-n*	*sug-i-den*	*käde-n*	*käz-i-den*
Partitive	*suga-d*	*sug-i-d*	*kä-t*	*käz-i-d*
Essive	*suga-n*	*sug-i-n*	*käde-n*	*käz-i-n*
Translative	*suga-ks*	*sug-i-kš*	*käde-ks*	*käz-i-kš*
Elative	*suga-späi*	*sug-i-špäi*	*käde-späi*	*käz-i-špäi*
Inessive	*suga-s*	*sug-i-š*	*käde-s*	*käz-i-š*
Illative	*suga-ha*	*sug-i-he*	*käde-he*	*käz-i-he*
Ablative	*suga-lpäi*	*sug-i-lpäi*	*käde-lpäi*	*käz-i-lpäi*
Adessive	*suga-l*	*sug-i-l*	*käde-l*	*käz-i-l*
Allative	*suga-le*	*sug-i-le*	*käde-le*	*käz-i-le*
Egressive	*suga-nnopäi*	*sug-i-dennopäi*	*käde-nnopäi*	*käz-i-nnopäi*
Approximative	*suga-nno*	*sug-i-denno*	*käde-nno*	*käz-i-nno*
Propinquative	*suga-nno(ks)*	*sug-i-denno(ks)*	*käde-nno(ks)*	*käz-i-nno(ks)*[109]
Abessive	*suga-ta*	*sug-i-ta*	*käde-ta*	*käz-i-ta*
Terminative	*suga-hasai*	*sug-i-hesai*	*käde-hesai*	*käz-i-hesai*
Prolative	*suga-dme*	*sug-i-dme*	*käde-dme*	*käz-i-dme*
Comitative	*suga-nke*	*sug-i-denke*	*käde-nke*	*käz-i-nke*

Table 3.16: Veps {SUGA} and {KÄZI} fully declined

The main functions of the case-forms in Veps are as follows:[110]

- nominative: the form of the telic subject of a finite verb and the direct object of an imperative;
- genitive: used adnominally to express possession and adverbally as the telic direct object; the form of the complement with most postpositions;
- partitive: used to express the atelic subject and atelic object; the form of the complement with most prepositions;
- essive (rare): (used only in the written language) encodes stative predicative adverbial functions;[111]

[109] "The forms with the *ks* cluster [...] are restricted to the easternmost [central] Veps dialects". Oinas (1961:131).
[110] Grünthal (2015a:63-97).
[111] Grünthal (2017) discusses the syntactic and morphological behaviours of the modern Veps essive in detail and surmises that, although marginally used, the case has not yet fully disappeared.

- translative: expresses a change of state;
- internal local case-forms: encode similar meanings as Finnish and Livonian (illative ‘into’; inessive ‘in’; elative ‘out of’);
- external local case-forms: used in similar ways as those found in Finnish: allative - ‘towards’, as well as the habitative functions ‘to’ and ‘for’; adessive ‘in, on, at’ as well as possession (e.g. *mölk vacas* ***kalal*** ‘the fish has roe in its stomach’), and the ‘agent’ or ‘instrument’; ablative - ‘away from’ and the habitative function ‘from’;
- ‘approximative’ local case-forms: includes the ‘propinquative’ (‘towards the vicinity of something/someone’), the approximative (‘in the vicinity or something/someone’), and the egressive (‘away from the vicinity of something/someone’);
- terminative: encodes the meaning ‘(right) up to’ both spatially and temporally;
- comitative: encodes the meaning ‘with’;
- abessive: expresses the caritive meaning ‘without’.

Neither consonant gradation nor vowel harmony is phonologically productive in Veps, although traces of the latter are found particularly in the illative singular case-forms (discussed later in this section). Inflexion classes in Veps are, as in other Finnic languages, generally identified on the basis of allomorphy in the stem. Differences between various stem allomorphs include: alternations to or additions of stem-final vowels (*suga*- and *robeh*-type nominals), alternations in the nature of a stem consonant (*käži*- and *puine*-nominals), or even metathesis of stem-final consonants (e.g., the nominative singular *lukkol* ‘lock’, genitive singular *luklo-n*). There are a few nominals which have invariant stems, such as nouns with a

monosyllabic nominative singular form of the shape *(C)V* (for instance, *ma* 'earth', *ö* 'night' etc.) and polysyllabic nominative singular forms ending in *-O* or *-U*.[112]

Zajceva (1981) offers the most thorough description of Veps inflexion classes,[113] identifying two major nominal groups: vowel-stems and consonant-stems. According to her analysis, vowel-stems can be found by removing the case exponent *-n* from the genitive singular form; consonant-stems can be found by removing the case exponent *-t* from the partitive singular form. This analysis suffers from a drawback, however, in that these two stem-types have to be found from different inflected forms. For instance, *avadim*-type nominals (table 3.17) should be considered consonant stems because removing the partitive singular exponent *-t* from the form *avadin-t* produces a stem which ends in a consonant, whilst removing the genitive singular exponent *-n* from the form *avadime-n* provides the speaker with what appears to be a vowel-stem. Zajceva's (1981) analysis of Veps inflexion classes may be broadly summarized as in table 3.17:

[112] In reference to Veps, a capital such as *-O* indicates that a vowel can be either front or back, but does not mean that the nature of the vowel is determined by vowel harmony.
[113] Zajceva (1995) offers a very similar analysis.

	Nominative stem	Common stem	Partitive stem	Illative stem	Plural stem	Gloss
Invariant stems (V)[114]			*ma*		*ma-i-*	earth
			*hebo**[115]		*hebo-i-*	horse
			longi		*long-i-*	dinner
Common stems in *-A* (V)			*kala**		*kalo-i-*	fish
			kova		*kov-i-*	hard
Common stems in *-e* (V)	*kivi*		*kive-*		*kiv-i-*	stone
	*sor'm**	*sor'me-*		*sor'm-*	*sor'm-i-*	finger
	avadim	*avadime-*	*avadin-*	*avadim(e)-*	*avadim-i-*	lock
	robeh	*robehe-*	*robeh-*	*robehe-*	*robeh-i-*	birchbark
	*käži**	*käde-*	*kä-*	*käde-*	*käž-i-*	cat
Common stems in *-iže-* (C)	*puine*	*puš-*	*puiže-*	*puiž-*	*puiž-i-*	wooden
	pakeine	*pakaš-*	*pakeiže-*		*pakeiž-i-*	frost
	edeline	*edeliš-*	*edeliže-*		*edeliž-i-*	former
Common stems in *-ha-* (C)	*veraz*	*verhas-*	*verha-*		*verh-i-*	stranger
	händikaz	*händikaha-*	*händikas-*	*händikaha-*	*händikazah-i-*	wolf
Common stems in *-se-* (C)	*eraz*	*erase-*	*eras-*	*erase-*	*eras-i-*	a certain
-ged > *-kt-* common stems (C)	*vouged*		*voukta-*		*voukt-i-*	white
tt-/-t't'- > *-čč-* common stems (C)	*kortte*	*korčče-*	*korttes-*	*korčče-*	*korčč-i-*	height
hn-/-zne-/-kne- > *-šš-* common stems (C)	*tohne*	*tošše-*	*tohiš-*	*tohišše-*	*tošš-i-*	birchbark
Metathesis common stems (C)	*lukkol*		*luklo-*		*luklo-i-*	lock
	ikun		*ikna-*		*ikno-i-*	window

Table 3.17: Veps inflexion classes[116]

[114] (V) indicates that the nominal type is a vowel-stem; (C) that it is a consonant-stem.

[115] Where marked with an asterisk, this nominal type also includes those with nominative stems ending in a consonant. The illative singular form vacillates in these nominal types between being characterized by distinct common and illative stem shapes, and being characterized by the same stem shape. For instance, in the *A*-stems, *kala* 'fish' has the illative singular *kalaha*, but *koir* 'dog' has an illative singular *koir-ha* (not **koira-ha*).

On consideration of the patterns of nominal inflexion exhibited by Veps, I identify five stem types:

- the 'nominative' stem - this stem is the same shape as the nominative singular and is sometimes also found in other inflected forms in the singular;
- the 'common' stem - the form found in most of the singular case-forms (although sometimes the nominative singular, partitive singular, and/or the illative singular forms are often characterized by distinct stem shapes) and the nominative plural;
- the 'partitive' stem - some nominals exhibit a stem shape in the partitive singular form which is distinct from all other stem allomorphs in the inflexional paradigm;
- the 'illative' stem - this is the stem that, in certain nominal types, is distinct from the nominative, partitive, and common stems and is found in the illative singular case-form;
- the 'plural' stem - this stem underlies all forms in the plural except for the nominative.

Only four nominal types exhibit five different stem shapes: *käži*-type nominals with a nominative singular form ending in a consonant (e.g., {PARZ'} 'lumber', {VOZ'} 'year'), and *avadim-*, *puine-*, and *vedetoi*-type nominals. For the most part, however, nominals do not exhibit five distinct stems. As in other Finnic varieties, the distribution of these stem shapes is consistent within the inflexion class, but not across different declensions; the distribution of stem allomorphs cannot be described in semantic or syntactic terms as is thus analysed here as morphomic.

---

[116] In addition to these inflexion types, two more declensional patterns that include mainly derived nominals may be identified: the caritive nominals ending in *-toi* (*vedetoi* 'without water', common stem *vedetoma-*, partitive stem *vedeton-*, illative stem *vedetom-*, plural stem *vedetom-i-*) and nominalizing forms in *-ut/-uz'* (*lapsut* 'childhood', common stem *lapsude-*, partitive stem *lapsu-*, illative stem *lapsude-*, plural stem *lapsuz-i-*). Zajceva (1981:213; 218), Zajceva (1995:71f.; 83f.). This table does not include very small classes (i.e., those with only one or two members).

Phonologically distinct case exponents across declension are more frequently found across (rather than within) the Veps dialects, particularly in the younger case-forms:

	Northern Veps	**Central Veps**	**Southern Veps**
Elative	*-spAi*	*-späi*	*-spää*
Ablative	*-lpäi*	*-lpäi/-pää*	*-upäi/lpää/-npää*
Terminative	*-hessai/hassai*	*-hussei/-hessei*	*-hessaa*
Prolative	--	*-dmu/-dme*	*-dmu/-mu*
Comitative	*-nke*	*-nke*	*-dmu/-mu*
Propinquative	*-nno(ks)*	*-nnost/-nno*	*-lost*
Approximative	*-nno*	*-nnou/-nno*	*-lon*
Egressive	*-nnopAi*	*-nnoupAi/-nnopäi*	*-lompä*

Table 3.18: Case suffixes in the three major Veps dialects

The variation exhibited by the forms given in table 3.18 is largely phonological and usually has to do with very minor differences in the quality of a vowel or diphthong. Only the means of expressing 'path' and 'accompaniment' differ significantly across the dialects, where in central and written Veps there are two separate morphological case markers for these two functions, but in southern Veps a single form *-dmu/-mu* expresses both 'path' and 'accompaniment', and in northern Veps the prolative is encoded by means of the postposition *möto*.

Within dialects, there is a great deal less formal variation in the inflexional suffix than in the stem. In written Veps there are two types of case suffix invariance: the nominative, genitive, translative, elative, and inessive are distinct only across number whilst all other case-forms are invariant across number and nominal type. The nominative singular, as in all other Finnic varieties, is always characterized by the lack of overt suffix and the nominative plural by final *-d*; the genitive singular by *-n* and the plural by *-den*. The translative, inessive, and elative forms exhibit differences between the singular and plural, where the preceding *-i-*

(usually the plural number marker, sometimes the stem-final vowel) has historically caused the *-s* to palatalize.[117]

Within individual dialects, the partitive singular, illative singular, and terminative singular are the only phonologically-distinct case exponents that are not conditioned by their phonological environment. First, in the written variety, depending on the nominal type, the partitive case-form can be either *-t* or *-d*.

Nom. sg.	***kaži***	***pen***	***se***	***vezi***	***vouged***	***vož***
Part. sing.	*kažid*	*pent*	*sed*	*vet*	*vouktad*	*vot*
Nom. pl.	*kažid*	*pened*	*sed*	*veded*	*vouktad*	*voded*
Gloss	cat	small	it	water	white	year

Table 3.19: Partitive exponents in Veps

Second, in the singular, there are eight desinences: *-he/-ha/-ho/-hu/-hä/-hö/-hy* and *-ze*; in the plural, only two: *-he* and *-že*. On consideration of certain illative singular forms, the exponent *-hV* may seem to be distributed according to the phonological environment. For instance, the forms in table 3.20 seem to follow a rule whereby the final vowel of the illative singular desinence *-hV* is a copy of the vowel immediately preceding the case suffix:

Illative singular desinence	**Nominative singular form**	**Illative singular form**	**Gloss**
-he	*kodi*	*kodi-he*	home
-ha	*čiža*	*čiža-ha*	older sister
-ho	*hebo*	*hebo-ho*	horse
-hu	*judu*	*judu-hu*	noise, racket
-hä	*hüvä*	*hüvä-hä*	good
-hö	*ö*	*ö-hö*	night
-hü	*idü*	*idü-hü*	nucleus, embryo[118]

Table 3.20: Distribution of illative singular exponents in Veps

[117] Hämäläinen and Andrejev (1934:12f.) claim that there exist Veps dialects in which only *-ks* is found for the translative case-form, although they do not specify which ones; Grünthal (2015a:75) offers an example from northern Veps in which the inessive form *-s* follows an *-i-*. The dialectal variation of these two case suffix allomorphs may well be more complicated than is suggested here, but in the written language the palatalization of the sibilant following *-i-* is absolute.

[118] In the last three examples, the illative singular allomorph can alternatively be characterized by the back equivalent can alternatively be found (i.e., *-a*, *-o*, and *-u*). The same does not hold of those illative singular desinences characterized by back vowel.

However, when a larger data set is considered, it becomes clear that the distribution of the different forms of the illative singular suffix are not always transparently determined by phonological considerations (as will be discussed in §4.3.2). Moreover, since the terminative singular has developed from the suffixation of a postposition **sāδen* which historically occurred with a nominal inflected in the illative, and since we have already claimed that the illative singular suffixal exponents cannot be described as being selected by the phonological environment, we may similarly conclude that the distribution of the terminative singular desinences cannot be described according to phonological conditioning environments.

3.3.2 THE DEVELOPMENT OF THE VEPS CASE SYSTEM - PHONOLOGY

In Veps, there are only rare instances of phonological developments leading to syncretism in the case system. One such example is the loss of final *-*t* at an earlier stage of the language which originally led to the elative and ablative becoming syncretic with the inessive and adessive case-forms respectively.[119] So, for instance, in a nominal such as {PERT'} 'house', the form **pertil* would have expressed both 'on/at the house' and 'away from the house'. In the modern dialects, however, the postpositional adverbial particle *päi* is found following the final *-l-/-s-* element.[120] As a result, the pattern of syncretism which was the result of phonetic erosion is no longer found in the modern dialects.

Furthermore, as a result of short final *-A*, the essive and the genitive values in the singular are now no longer encoded by distinct morphological forms. This is a rare instance in which phonological changes have led to the near-total loss of a case value, though only in the singular, as the essive and genitive forms in the plural would have remained distinct.

[119] Tikka (1992:22), Grünthal (2003).

[120] The semantics of the 'younger' source case-forms do not exactly match those of the 'older' source forms. The external local cases in the Finnic languages tend to govern more grammatical uses such as 'possessor' and 'indirect object'. The internal local cases are also used to express non-local meanings. However, the synchronic elative and ablative (and egressive) have a semantic constraint [+spatial], which means that, for instance, the ablative is used to express 'place from where' but not habitative functions. Again, therefore, we find functional shifts alongside other developments.

Although Grünthal (2017) argues that there is still evidence to suggest that the essive may be analysed as morphologically and semantically distinct from the genitive, it cannot be considered a productive value in the same way as, for instance, the local case-forms. It is possible that phonetic erosion led first to the loss of a phonological distinction between the essive and genitive in the singular, and that thereafter there was a gradual loss of the functional distinction once maintained by the morphological difference. Phonetic erosion and functional opacification thus both played a part in the loss of the essive case value in Veps.

Phonetic erosion has also had an effect on the postpositional material which has grammaticalized into new case suffixes. In Veps, as many as eight new case-forms have developed (or, more accurately, are in the process of developing) from postpositions and postpositional adverbs. The postposition **lō-* (cf. Finnish *luo(kse)* 'towards', *luona* 'at, near', *luota* 'from'), for instance, is now generally analysed as a case suffix in all the major Veps dialects.[121] In the majority of instances, this postposition has fused onto the genitive form of the preceding nominal and the initial labial of the postposition has assimilated to the genitive singular exponent *-n*, e.g., *-n* + **lō-* > **-nlō* > *-nno* in central and northern Veps.[122] In southern Veps, the genitive *-n* is not preserved and the *-l-* of the postpositional form remains.

Whilst it is generally agreed that the approximative local case-forms have their origins in an inflected form of the postposition **lō-*,[123] it is not always agreed from precisely what form(s) of **lō-* they have developed from. The approximative local case-forms in table 3.16 present a misleadingly simple account of the data, and gloss over a great deal of the formal variation found across the three major Veps dialects:

[121] Certain Lude and Karelian dialects have also developed new 'approximative' case-forms from the suffixation of the postposition **lō-*. Oinas (1961:137).
[122] Tikka (1992:47).
[123] Oinas (1961:108), Tikka (1992:54), and Grünthal (2003:123).

	Written Veps	Northern Veps	Central Veps	Southern Veps
Propinquative	*-noks*	*-nou, -no, -nol*	*-nost; -noks, -loks*	*-lost*
Approximative	*-no*	*-no, -nu*	*-nou, -no, -nol*	*-lon*
Egressive	*-nopäi*	*-nopAi*	*-noupAi*	*-lompā̄*

Table 3.21: Approximative local case-forms across the Veps dialects[124]

Of particular interest is the propinquative, the origins of which have been widely debated. The presence of the final element *-ks* in the easternmost dialects of central Veps contrasts with the final consonant cluster *-st* in other central Veps dialects and in southern Veps. Forms in *-st* are usually reconstructed as developing from **lōsten*, which is analysed as the form **lō-*, a lative element *-s-*, and a particle suffix *-ten*, whilst forms in *-ks* are hypothesized to originate from a lative form **lōksek*, a cognate of Finnish *luokse* 'to, towards'. It has been argued that a development **loks* > **lost* can account for the two distinct forms across the dialects,[125] although this is contested by Oinas on the grounds that the same phonological development did not occur elsewhere in the language.

In northern Veps, moreover, we find a propinquative form which is identical to the approximative and is not characterized by either final *-ks* or *-st*. A strictly phonological approach struggles to derive forms such as *-nou, -no,* and *-nol* from either **lōsten* or **lōksek* and, instead, Tikka suggests that these suffixes may have developed from two distinct forms, **lōk* (for northern Veps) and **lōksek* (central and southern Veps).[126] This preserves a phonological analysis, although it does potentially raise a methodological issue, namely: where a phonological analysis is problematic (particularly where data are lacking), what is to stop us from 'saving' the analysis by simply stating that desinences were not derived from a single historical form, but from two or more distinct forms? For instance, rather than posit the

[124] Adapted from Tikka (1992:47). Even this glosses over variation found within these dialects. Oinas (1961:137) and Kettunen (1943:369f.).
[125] Kettunen (1943:370).
[126] 1992:134.

contentious change **ks* > **st*, could we not simply argue that there were two forms **lōksek* and **lōsten* which existed, and perhaps competed, at an earlier stage of Veps, and where one form 'won out' over the other in particular dialects, whilst the second remained elsewhere? This hypothetical question is not necessarily intended to seriously suggest that the propinquative forms *-ks* and *-st* originate from two separate words: further research would be required to determine whether such a hypothesis were valid or even worth pursuing. Instead, it is meant to highlight the weaknesses inherent in relying solely on phonology as an explanatory factor in the development of a given language, and particularly one with a very short history of attestation, such as Veps.

Finally, in the southern dialect, forms ending in *-dmu* is used for prolative, comitative, and instrumental functions. This combination of functions is not encoded by a single inflexional form anywhere else in the Finnic space. Even elsewhere in Veps (as in table 3.18), prolative and comitative/instrumental functions are expressed by separate forms. Kettunen asserts that there is nothing remarkable in the prolative developing comitative meanings in southern Veps.[127] However, there are no other instances in the Finnic sphere of a single form encoding both 'prolative' and 'comitative/instrumental' functions. Instead, it might be suggested that the prolative-comitative suffix in southern Veps is the result of the suffixation of a postposition which was already polyvalent as, in Finnish, there is a postposition *myöten*, which encodes prolative functions, as well as a postposition/adverb *myötä* (sharing a root *myö-*), which can express a comitative meaning. The form *-mu* found in the southern Veps has its origins in a postposition **möto*, related to these Finnish forms, and so it might be possible to posit a development whereby two originally different postpositional forms 'merged' phonologically and functionally *before* grammaticalizing into an inflexional case-

[127] 1943:362, 545.

form in southern Veps.[128] This view is problematized, however, by consideration of northern and central Veps, in which comitative and instrumental functions are encoded by distinct forms.

Although phonological developments can account in some part for the shapes of some of the inflexional case suffixes found in Veps, it is not clear that they can offer satisfactory explanations for all the changes which have affected the case system, and particularly the impressive innovation of novel inflexional case-forms. Without further evidence of an earlier stage of Veps, of course, it is not possible to answer for certain many of the questions that have been raised here.

3.3.3 THE DEVELOPMENT OF THE VEPS CASE SYSTEM - ANALOGY

It is much more difficult to detect analogical developments - either those which have already affected the language or those which are ongoing - in the Veps language than it is in Finnish. The main reason for this is the lack of resources available which discuss such developments (particularly in reference to the case system). That is not to say that there are not plenty of analogical developments ongoing in the modern language, just that, given the evidence available, it is difficult to detect them.

In Veps there appears to be an analogical development of a distinct case-form on the model of the implicational relations that exist between sub-word elements.[129] There are, for instance, a number of arguments to suggest that the development of the source case-forms in *-päi* is analogical, rather than strictly phonological, in nature. In the first place, the grammaticalization of the source forms in *-päi* is not yet complete. Sometimes morphological stasis forms may be used to encode particularly the habitative functions that might once have

[128] My thanks to Ulla-Maija Forsberg (p.c.) for bringing this suggestion to my attention.
[129] cf. the discussion on the Finnish excessive in §3.1.3.

been expressed by the etymological source forms.[130] Furthermore, the element *-päi* is sometimes separated from the nominal stem by means of possessive suffixes: *tat-s-eiń-pei tuli-ń* father-EL.[1]-1POSS.SING.-EL.[2] come-IMPF.-SING.1 'I come from my father's (home)',[131] which indicates that it still exhibits some of the behaviours of a postpositional adverbial particle.

Whilst one of the source forms (the ablative, elative, or egressive) may have developed from the blurring of the prosodic boundary between the case element *-s-*, *-l-*, or *-(n)no-* and the particle *päi*, the spread of what was perceived as a 'source' marker may have been at least partly analogical, rather than purely phonological, in nature. Unfortunately, such a suggestion remains speculative, since the written evidence is of such a restricted nature that the principled interrogation of the development of the suffixation of *-päi* is very difficult. Nevertheless, in a similar way to the development of the Finnish excessive discussed in §3.1.3, the sub-word clarity of forms in which *-päi* was beginning to be analysed as part of the morphological word may have aided its generalization not only to other lexemes but to other local case series (e.g., from the external to the internal, or from the internal from the approximative, etc.). Thus, the patterns - not only formal, but also functional - which existed between the sub-suffixal elements of the source and stasis forms may have been extended throughout the language as a direct result of their phonological clarity and semantic relatedness.

One further example of what appears to be analogical change is the occasional fluctuation of the partitive exponents *-d/-t*. In lexemes such as {AVADIM} 'key' and {SAVI} 'clay', there seem to be two partitive singular forms, one ending in *-t* and the other in *-d*: *avadime-d* and

[130] Grünthal (2003:124).
[131] Zajceva (1981:188f.).

avadin-t; *save-d* and *savt'*.[132] Although nowhere can I find discussion on this apparent overabundance (the two forms *avadimed* and *avadint* are given in different resources: the former is much more common; the latter I have found only in the Veps~Latvian dictionary),[133] that two forms are given may suggest that they are 'acceptable' to some degree. If these forms are both found in a single dialect, then it is possible that the irregular form in *-t* is being ousted by the form *-d* (which is much more common across declension classes and lexemes: *-t* tends to remain in certain more frequent lexical items) by means of analogy. However, since I am aware of no discussion on the subject, it is possible that the different forms are actually found in different dialects, and that they reflect slightly different phonological developments, rather than the analogical spread of the partitive form *-d*.

3.3.4 THE DEVELOPMENT OF THE VEPS CASE SYSTEM - SYNTAX

There are a number of 'younger' or 'newer' case-forms in Veps which have developed from postpositions or postpositional adverbial particles. These behave in slightly different ways from the 'older' case-forms in terms of a) whether they licence concord, and b) whether they occur with the common stem or an inflected form of the lexeme. It is not necessarily easy to determine whether the newer case-forms in Veps are 'fully-fledged' morphological case-forms or whether they are still better described as postpositions.

[132] Andrejev and Hämäläinen (1934:17).
[133] Rozentāls (2008:16).

	Modern forms	Postposition/postpositional adverb
Comitative	*-nke*	*[[N + GEN] + kera]*
Prolative	*-dme*	*[[N + PART] + möto*
Ablative	*-lpäi*	*[[N + ADES] + päi]*
Elative	*-späi*	*[[N + INES] + päi]*
Egressive	*-nnopäi*	*[[N + GEN] + lōpäi]*
Approximative	*-nno*	*[[N + GEN] + lō]*
Propinquative	*-nnoks*	*[[N + GEN] + lōksek]*[134]
Terminative	*-hVsai*	*[[N + ILL] + sāδen*

Table 3.22: The development of new inflexional case-forms in Veps

As was mentioned in §1.1.2, it is not always clear where to draw the line between 'case' and 'adposition', although canonical generalizations may be appealed to in order to highlight the differences between the two. The canonical case will licence concord (meaning that all adjectives agree with their head noun in both number and case). Furthermore, lexical material cannot 'interrupt' an inflected form (that is, it cannot occur between the lexical stem and the inflexional material). In Veps, the case-marker never takes main stress, unlike postpositions. A case-suffix will occur with the lexical stem (although the shape of the stem may differ, as discussed in §3.3.1), whilst a postposition takes a nominal complement inflected in either the genitive or the partitive case; postpositional adverbial particles such as *päi* co-occur with a nominal inflected in a local case such as the inessive or adessive. Finally, although not an instance of canonical case structure, in Veps there is a tendency for inflexional case exponents to be shorter (usually one to two syllables) than postpositions. According to these criteria, it is clear that the younger morphological case-forms do not always behave like the older (and frequently canonical) case-forms:

[134] cf. the discussion on the development of the approximative local case-forms in §3.3.2.

Form	Licenses concord	Attaches to nominal stem	Cannot be separated from nominal stem	Cannot take main stress	(Two syllables or fewer)
Genitive singular *-n*	X	X	X	X	(X)
Written Veps Comitative *-nke*	-	-	X	X	(X)
Written Veps Approximative *-nno*	-	-	X	X	(X)
Southern Veps approximative *-lo* (discussed below)	X	X	X	X	(X)
Postposition *tagapäi* 'from behind'	-	-	-	-	(-)

Table 3.23: Comparing morphological case-forms and postpositions in Veps

The genitive singular case-form *-n* is unproblematically analysed as an inflexional suffix; the postposition *tagapäi*, a postposition. The written Veps comitative and approximative forms (*-nke* and *-nno* respectively), however, exhibit mixed behaviours. Moreover, as can be seen from table 3.23, the status of the newer case-forms is not uniform across the three dialects and written variety. In southern Veps, for instance, the approximative local case-forms are starting to more consistently licence concord and do not occur with a nominal inflected in the genitive. These younger suffixes in this dialect are therefore further along the cline of grammaticalization, and it is easier to describe them as inflexional case suffixes than the cognate case-forms in the other dialects.

It is not immediately clear, however, why Veps has developed so many new case-forms from postpositions and other postpositional material whilst Finnish exhibits only one new form from a postposition. It cannot simply be claimed that Veps propagated its case system *because* it is characterized by postpositions, as the same development is not found in either

Finnish or Livonian, despite the presence of postpositions in these Finnic varieties. This will be discussed further in §6.4.

3.3.5 THE DEVELOPMENT OF THE VEPS CASE SYSTEM - CONTACT CONSIDERATIONS

As was mentioned in §3.2.5, the sociolinguistic situation in which Livonian was spoken in its last years shares many similarities with Veps. Veps, too, has been in long-term and intense contact, but with Russian (as opposed to Latvian), and the language and its users have not always had the same rights as Russian and its native speakers. For instance, the development of a written variety of Veps occurred relatively late, and it was used for only a few years - between 1934 and 1937 - at which point "all cultural activities were forbidden, Veps[...]-language schools were shut, Veps[...] books were destroyed and teachers deported or shot."[135] Speakers of Veps were scattered first of all by the government and later by the Second World War. Although the situation is less bleak today, Veps is not learned natively by children and tends to be spoken in very restricted environments - i.e., at home with other native speakers of Veps, who are usually over 65.[136] Older native speakers of Veps do not tend to consider it important for younger people to learn the language, although the younger generations are keen for children to acquire the language as part of their heritage.[137]

All native speakers of Veps are, by necessity, at least proficient in Russian; the younger generation is usually fluent. According to the work conducted by scholars such as Thomason and Kaufman (1988) and Trudgill (2011, this sort of bilingualism (that in which the two languages are natively learned) can sometimes lead to 'complexification'. Although in Veps we do find a large number of loanwords from Russian, howeover, it is less clear that such intense language contact has directly caused or contributed to any developments found in the Veps case system. One counter-example may be found in the development of the southern

[135] Blokland and Hasselblatt (2003:128).
[136] Puura et al. (2013:50).
[137] Blokland and Hasselblatt (2003:129).

Veps prolative-comitative form, which, as was discussed in §3.3.2, cannot easily be analysed as a change which has occurred solely as the result of regular phonological changes. Grünthal suggests instead that language contact may have exerted an influence on the development of this typologically odd (at least in the Finnic sphere) property of the Veps case system.[138] That is:

- the southern Veps prolative-comitative form *-mu* developed from the suffixation and grammaticalization of a postposition **möto* which encoded prolative functions only;
- Veps speakers were familiar with Russian, in which there is an instrumental case which encodes not only 'instrumental,' but also 'prolative' and 'comitative', functions;
- recognizing the tripartite use of the Russian instrumental case, speakers of Veps analogically extended comitative (and instrumental) functions to a form which originally expressed only path.

The main drawback with this analysis is that appealing to language contact may work relatively well, except that all the major Veps varieties are in intensive contact with Russian. Why, therefore, do we not find the same form encoding prolative and comitative functions in central or northern Veps as well as the southern dialect? Without further information about the contact situation of Veps through time, we may be in danger of referring to contact as a causal factor in certain developments simply because they cannot otherwise be easily explained.

The weaknesses of appealing to contact as an explanatory factor in language development can also be seen by directly comparing the situations in which Veps and Livonian have been

[138] 2003:165-177.

used and by noting the completely opposite developments of their case systems - that is, the accretion of inflexional case-marking in Veps and the loss of them in Livonian. By appealing to contact alone, it is difficult to account for the very different changes we find in Livonian and Veps when they appear to have been spoken in such similar situations for at least the last hundred years:

Livonian	**Veps**
No longer natively learned	Rarely natively learned
In long-term and intense contact with a typologically very different language - Latvian	In long-term and intense contact with a typologically very different language - Russian
In long-term and intense contact with a language which maintains a smaller inflexional case system - Latvian	In long-term and intense contact with a language which maintains a smaller inflexional case system - Russian
Culture and language suppressed for many years before being officially recognized and supported	Culture and language suppressed for many years before being officially recognized and supported

Table 3.24: The similarities between the Livonian and Veps sociolinguistic situation

The main difference between these two Finnic varieties appears to be the fact that Veps is in near contact with another Finnic language, Karelian.[139] The Livonian dialects spoken in Courland are separated from their nearest Finnic neighbour - Estonian - by the Baltic sea, or by a rather circuitous route by land. Since Livonian is slightly more cut off from contact with related varieties which maintain predominantly agglutinative structure, it may be that this facilitated the morphological typological shift from predominantly agglutinative to predominantly fusional, particularly since it was in such intense and prolonged contact with the fusional language Latvian. Veps, on the other hand, was still in contact with other typologically similar languages, and this may have had facilitated the retention of the inflexional case system. Such suggestions must, however, remain speculative.

[139] Furthermore, if Lude, although usually considered a dialect someway between Veps and Karelian, is counted as a separate Finnic language, then Veps can be considered to have been in long term contact with this Finnic variety, too.

3.4 CLOSING REMARKS

The developments of the case systems of the Finnic languages discussed here are in a number of ways more disparate than that of the Italic varieties outlined in chapter two. There appears to be one example of an unequivocally 'case-reducing' language (Livonian) and one example of an unequivocally 'case-increasing' language (Veps, particularly the southern variety in which the younger case-forms have grammaticalized to a greater extent). Finnish, on the other hand, is more difficult to analyse. It appears to be in the process of losing at least four case values - the abessive, comitative, prolative, and instrumental - which are in restricted usage in the modern language, whilst there is also evidence to suggest that at least one new case-form in the process of developing (the comitative), and another which was at an earlier stage of the language in development (the excessive) but which is now restricted largely to particular lexemes in particular dialects. Rather than uncontroversially case-increasing or case-decreasing, I analyse Finnish as a 'case-steady' language (§1.0) since although exhibiting fluctuations in the statuses of certain case-forms (i.e., whether they have fully grammaticalized or whether they are stylistically or regionally restricted), the 'core' values of the case system have remained unaffected by significant accretion or depletion through time.

Although certain developments found in the Finnic languages under consideration here may be explained with relative ease as being the results of phonological, analogical, syntactic, or contact-induced changes, it is not clear why we find such different developments across varieties. Nor is it clear why the development particularly of the Finnish and Veps case systems differs so drastically to those of the Italic languages discussed in chapter two. The rest of this study is concerned with analysing the data presented in here and in chapter two and offering observations about what motivates the loss, maintenance, and accretion of case-forms through time. In order to do this, it is necessary first to identify and describe some of

the morphological structures that are found in these typologically quite distinct varieties, as it will be later suggested that the different morphological structures found in language have an effect on: a) how that language variety is acquired and used by learners, and b) how the language is likely to develop in respect to its case system.

4.

THEORETICAL APPROACHES TO THE ITALIC AND FINNIC DATA

4.0 INTRODUCTION

This chapter offers an analysis of the Italic and Finnic data presented in the previous two chapters from both *constructivist* and *abstractivist* perspectives (the basic principles of which will be outlined in §4.1).[1] Whilst it will be concluded that an abstractivist model best accommodates the morphological structures found in the varieties under discussion in this study, a constructivist (particularly a 'morphemic') approach more clearly highlights a significant difference between the syntagmatic structure of inflected wordforms in the languages introduced in chapters two and three. That is: whilst Livonian and all of the Italic languages discussed here exhibit a great deal of fusional structure in their case systems, Veps and Finnish are characterized by an abundance of agglutinative structure, in which form and function are in a transparent relationship. Work conducted from an abstractivist view does not usually recognize a qualitative difference between morphological structure in which the form~function mapping is direct and that in which it is 'interrupted' (discussed particularly in reference to the Italic languages). However, it can capture the different ways in which these morphological structures are acquired and utilized in the production of novel inflected forms.[2] This, in turn, sheds light on why languages tend to lose, maintain, or propagate inflexional case-marking, as will be explained in greater detail in chapters five and six.

4.1.0 MODELS OF MORPHOLOGICAL DESCRIPTION

Although there are many different models of morphological description, there are similarities in the ways that these approaches hypothesize that surface inflected forms are generated. Stump (2001) splits theories of inflexional morphology into four different types,

[1] The terms 'constructivist' and 'abstractivist' are adapted from Blevins' (2006) terms 'constructive' and 'abstractive'.
[2] The ways in which inflected forms are processed is not discussed in this study.

based on how the morphosyntactic properties of an inflected form are associated with particular aspects of its morphological shape and whether inflexional morphology is information-increasing or associative:

	Lexical	**Inferential**
Incremental	Lexical-Incremental	Inferential-Incremental
Realizational	Lexical-Realizational	Inferential-Realizational

Table 4.1: Theories of inflexional morphology

In lexical theories, the associations between a form and its morphosyntactic properties are listed in the lexicon. For instance, the suffix *-s* in English has a lexical entry which specifies its association with the morphosyntactic property 'plural'. Inferential theories, on the other hand, make use of rules or formulae which are used to express the systematic formal relations that exist between a lexical root (or stem) and the fully-inflected wordform. In incremental theories, morphosyntactic information is added by means of meaning-bearing inflexional exponents (and may therefore be considered 'information-increasing'). Finally, realizational theories are those in which "a word's association with a particular set of morphosyntactic properties licenses the introduction of those properties' inflectional exponents".[3]

In contrast, Blevins (2006) restricts his focus to the different groups of mutually consistent assumptions associated with approaches to morphological description and identifies only two different model types which are here termed 'constructivist' and 'abstractivist'.[4] Since this distinction subsumes all known models of grammatical description, including those that can be encompassed within Stump's distinctions, I will offer a more in-depth introduction to these two different approaches here.

[3] Stump (2001:2).
[4] From Blevins' (2006) terms 'constructive' and 'abstractive'.

4.1.1 ‘CONSTRUCTIVIST’ APPROACHES

In constructivist models, inflected words are ‘built up’ or ‘constructed’ from an inventory of lexical stems and either inflexional exponents or inflexional processes. In one such model, Item and Arrangement (IA), forms are built up by the concatenation of minimal, meaning-bearing units, or ‘morphemes’. According to this approach, the plural of the lexeme {CAT} is produced by concatenating the lexical stem /kæt/ with the plural morph /s/ to give the surface form /kæts/. Inflexional functions are introduced by means of meaning-bearing units which are agglutinatively concatenated onto an invariant lexical stem in a biunique mapping of function and form. IA may be described as a ‘lexical-incremental’ theory according to Stump’s (2001) terminology.

In an Item and Process (IP) model, on the other hand, inflected forms are built from the concatenation of a lexical stem and ‘process’ morphs. The latter are units that indicate a word form has undergone certain processes that associate a given form with particular morphosyntactic properties, rather than phonological elements which *bear* meaning (as in IA). As Hockett puts it: “[s]ome of the phonemic material in a derived [i.e., inflected] form may be, not part of any underlying form, but rather a representation or marker of the process.”[5] These process morphs can be used to account for certain changes in form which distinguish functions but which are not necessarily introduced by overt affixal material. For instance, the English form /fi:t/ is the plural form of lexeme {FOOT}; here the difference in number is not encoded by an affix but by ablaut (/ʊ/ > /i:/). In IP, a zero-morph -*Ø* expressing plural functions may be posited which would affect the shape of the lexeme stem. This model may be described as ‘inferential-incremental’ in Stump’s terminology. Since both IA and IP approaches prioritize sub-word, meaning-bearing forms as the smallest units of analysis, these models are often referred to as ‘morphemic’ models.

[5] 1954:396. Emphasis original.

In morphemic approaches both words and paradigms as theoretical constructs are regarded as epiphenomena. When speakers 'build' inflected forms, models such as IA and IP do not allow for reference to be made to other inflected forms of a lexeme or to the distribution of stems, inflexional exponents, or whole wordforms in the inflexional paradigm.

Although models such as IA and IP differ in Stump's terminology, according to Blevins' analysis of morphological models of grammatical description, they both constitute examples of *constructivist* perspectives, since the underlying assumption is that fully-inflected forms are built up of sub-word units. Not all constructivist models posit the morpheme as the minimal unit of analysis, however; nor do they all regard the paradigm as an epiphenomenon. Approaches such as A-Morphous Morphology (Anderson 1992) and Paradigm Function Morphology (Stump 2001), for instance, are better described as 'stem-' or 'root-based', and the latter in particular affords central importance to the paradigm for the realization of inflected forms.[6] In these approaches, sub-word parts are not necessarily considered 'meaningful' in isolation, and fully-inflected forms are constructed generally by '(realization) rules' or 'formulae'.

All of these constructivist models are characterized by similar weaknesses which are inherent in the recognition of minimal forms from which larger units are constructed or derived. For instance, positing 'stems' or 'roots' as minimal in these approaches necessitates the identification of 'stems' or 'roots' (as opposed to 'inflexional material' such as affixes), which is not always straight-forward. Since 'stem-'/'root-'based approaches both require the decomposition of fully-inflected forms, they suffer from the same drawbacks as will be discussed in §4.2.1 and §4.3.1 in reference to morphemic approaches. 'Stem-'/'root-'based approaches will therefore not be considered in as much depth as IA and IP in this chapter.

[6] More recently, Ackerman et al. (to appear) have suggested a slight re-work of Paradigm Function Morphology to make it (more) abstractivist rather than constructivist in nature. Since this represents only an initial attempt, the model is still analysed as constructivist here.

The reader is referred to Anderson (1992) for details about A-Morphous Morphology, to Stump (2001) for Paradigm Function Morphology, and to Blevins (2006) and (2016:119-162) for more information on the problems associated with these models.

4.1.2 'ABSTRACTIVIST' APPROACHES

In contrast to constructivist perspectives, an abstractivist approach such as Word and Paradigm (WP) does not posit that wordforms are constructed from the combination - whether by concatenation or by realization rules - of sub-word units.[7] Recurrent partials can in certain instances be abstracted away from fully-inflected forms and may then be used by speakers in the inflexion of novel forms, but forms are not considered to be 'built' in isolation; instead, it is recognized that "[t]he *word* is a more stable and solid focus of grammatical relations than the component morpheme by itself."[8] Implicational relations that exist between multiple fully-inflected wordforms provide information about the production of novel inflected forms, since certain morphosyntactic properties can be associated with particular forms only by reference to other inflected forms in the paradigm.

Despite the apparent priority afforded to 'word-'level units, it is recognized that even these are abstractions rather than persistent units (though, as is indicated in the quotation from Robins above, it is often suggested that they are abstractions that tend to provide greater certainty for speakers than sub-word elements such as morphemes or stems). Abstractivist approaches are therefore *unit-agnostic*. That is, although 'words' are often afforded a prioritized place in this perspective, sub-word units and periphrastic constructions, for instance, may also be abstracted away from the speech stream, and the implicational patterns

[7] A more apt name for this abstractivist approach is 'Item and Pattern' which captures its unit-agnostic nature (discussed later in this section); the name 'Word and Paradigm' prioritizes word-level forms and paradigmatic patterns over all others. In order to avoid confusion, I will not refer to 'IP', since this shorthand is here used instead for the 'Item and Process' model. However, it should be stressed that the WP approach described here recognizes *all* units and patterns that may be informative about the production of novel forms, not just words and paradigms.

[8] Robins (1959:128). Emphasis added.

that exist between these forms may be just as informative, in certain instances, as the abstract relations that exist between 'words'. Stems and inflexional exponents can be described as 'units of analysis' where they provide a language-user with information about the inflexion of novel lexemes. However, in WP they are not assigned meaning in isolation as in many constructivist accounts.

Importantly, it is not just fully-inflected wordforms (sub-word units, periphrastic constructions, etc.) themselves that provide the speaker with essential information about other inflected forms. Instead, it is the wordform *in addition to the function(s) it encodes* which provides the relevant information about inflexional patterns a nominal can follow. For instance, in the following proportional analogies in Latin, the strictly formal parallelism of three wordforms leads to the deduction of an incorrect genitive singular form for the lexeme {TEMPLA}:

PUELLA : PUELLAE 'girl'

MENSA : ? 'table'

? : MENSAE

But:

PUELLA : PUELLAE 'girl'

TEMPLA : ? 'temple'

? : ***templae*

Figure 4.1: Purely formal four-part analogy of Latin {TEMPLUM} 'temple'

In the first example, PUELLA is the nominative singular form and is characterized by a form ending in *-a*; the genitive singular form ends in *-ae*. MENSA 'table' exhibits a similar

structure to PUELLA in that it ends in *-a*, and on consideration of this surface form, we may suggest that the unknown '?' form is MENSAE - which is, for this lexeme, the correct form. However, on consideration of form alone, the form TEMPLA, which, like PUELLA, ends in *-a*, may lead us to suggest that the target form is ***templae*, one not found in this lexeme's inflexional paradigm. Instead, knowing that the form TEMPLA encodes nominative/accusative plural functions rather than nominative singular functions permits the speaker to determine that this lexeme does not follow the same inflexional pattern as {PUELLA} and {MENSA}. Throughout, I will refer largely to the forms of inflected nominals, but it should be understood that this refers implicitly to the form-function combination.

Another key notion in an abstractivist approach is that of 'entropy', the measure of uncertainty associated with an event. The following formula is used to calculate the amount of uncertainty in *X*:

$$H(X) = -\sum_{i} p(x_i) \log_2 p(x_i)$$

Figure 4.2: Calculating entropy

Here, "[t]he entropy $H(X)$ is the weighted average of the *surprisal*-$log_2 P(x_i)$ for each possible outcome x_i," calculated in bits.[9] Let us take a coin toss as a simple example. There are two outcomes for a coin toss - heads or tails - each of which is equally likely (as long as the coin is not interfered with in any way). We can calculate the uncertainty of either of the outcomes 'heads' or 'tails' in terms of bits using the entropy calculation given in figure 4.2, where the result of '1' in figure 4.3 is the "is the information in a choice between two equally probably outcomes."[10] This means that each of the results is equally probable. Where a result is certain (for example, if a coin were rigged to always land on heads), the probability is 0

[9] Ackerman and Malouf (2013:438).
[10] Ackerman and Malouf (2013:438).

bits since the outcome can be predicted in every instance. The more uncertain the result, the higher the number of bits.

$$H(X) = -\sum_{i}(P(x_i)\log_2 P(x_i)$$

$$= -(P(heads) * \log_2 P(heads) + P(tails) * \log_2 P(tails))$$

$$= -(0.5 * \log_2 0.5 + 0.5 * \log_2 0.5)$$

$$= 1(\text{bit})$$

Figure 4.3: Calculating the entropy of a coin toss

'Conditional entropy' measures the information provided by two or more wordforms in an inflexional paradigm, and thus provides a means of calculating how predictable (and predictive) certain wordforms are. The lower the entropy of a given cell, the more predictable the form in that cell is; the higher the entropy measure, the less predictable that form is.

$$H(\mathrm{C}_2|\mathrm{C}_1) = -\sum_{x \in Rc_1}\sum_{y \in Rc_2} p(x,y)\log_2 p(y|x)$$

Figure 4.4: Calculating conditional entropy

Conditional entropy calculates how much the uncertainty associated with producing a novel form is reduced by joint knowledge of one (or more) inflected wordforms. The measure in bits is that of the uncertainty (entropy) that remains associated with deriving the unknown form given knowledge of another.[11]

Utilizing the sorts of abstract patterns that can provide information in the deduction of a novel form is not always a maximally-predictive means of producing a previously-unencountered inflected item. The inflexional case suffixes of Modern Greek demonstrate this point very clearly:

[11] Blevins (2016:176).

	Singular			Plural		
Class	**Nom.**	**Gen.**	**Acc.**	**Nom.**	**Gen.**	**Acc.**
1	*-os*	*-u*	*-on*	*-i*	*-on*	*-i*
2	*-s*	*- Ø*	*-Ø*	*-es*	*-on*	*-es*
3	*-Ø*	*-s*	*-Ø*	*-es*	*-on*	*-es*
4	*-Ø*	*-s*	*- Ø*	*-is*	*-on*	*-is*
5	*-o*	*-u*	*-o*	*-a*	*-on*	*-a*
6	*- Ø*	*-u*	*-Ø*	*-a*	*-on*	*-a*
7	*-os*	*-us*	*-os*	*-i*	*-on*	*-i*
8	*- Ø*	*-os*	*-Ø*	*-a*	*-on*	*-a*

Table 4.2: Modern Greek nominal inflexion classes[12]

We may calculate the information still required to predict a specific unknown inflected form from another by means of conditional entropy $H(c_2|c_1)$. The conditional entropies for the Modern Greek inflexional case-forms are as follows:

H(*row*¦*column*)	NOM.SING.	GEN.SING.	ACC.SING.	NOM.PL.	GEN.PL.	ACC.PL.	*E*[row]
NOM.SING.	--	1.000	0.250	0.750	0.000	1.000	0.571
GEN.SING.	0.594	--	0.594	0.594	0.000	0.594	0.509
ACC.SING.	0.451	1.201	--	0.951	0.000	0.951	0.644
NOM.PL.	0.594	0.844	0.594	--	0.000	0.250	0.411
GEN.PL.	1.750	2.156	1.549	1.906	--	2.156	1.853
ACC.PL.	0.594	0.594	0.344	0.000	0.000	--	0.268
E[column]	0.719	1.120	1.120	0.736	0.000	0.879	0.664

Table 4.3: Conditional entropies for Modern Greek paradigms in table 4.2[13]

According to these conditional entropies, in Modern Greek knowledge of the genitive singular form gives a language-user enough information to select the correct nominative singular form from two equally probable options. In many instances, knowledge of one other form in the paradigm is sufficient to decide between those two choices. However, in some instances (such as predicting the genitive singular form from the genitive plural form), even more information is required to be able to deduce the correct shape of the inflected lexeme. Only the genitive plural exponent, since it is invariant across declension, can be predicted

[12] The vocative is not included in tables 4.2 and 4.3.
[13] Tables 4.2 and 4.3 adapted from Ackerman and Malouf (2013:434 and 441 respectively).

with certainty. However, as a result of its invariance across declension, its predictiveness is very low.

In this study, I will make reference to 'minimally-entropic' (i.e., 'maximally predictable and predictive') material, particularly in reference to suffixal morphology in the Finnic languages. These are forms which are phonologically invariant and which can be consistently associated with a particular function, and therefore afford a speaker a great deal of predictability in the production of novel forms.

The notion of entropy is particularly informative when it comes to answering the so-called 'Paradigm Cell-Filling Problem': given exposure to a novel form, how does a language-user fill other cells in the lexeme's paradigm?[14] In IA and IP, language-users supposedly have knowledge of sub-word units and processes which they may, theoretically, concatenate onto new lexical stems to which the hearer is exposed in order to produce novel inflected forms. In WP, on the other hand, it is the abstract relations that exist between fully-inflected wordforms which are used in the production of novel forms. The notion of entropy provides WP with a means of measuring the informativeness of particular forms and the implicational patterns that exist between them.

An abstractivist approach is implicit in the 'classical' or 'traditional' didactic grammars of languages such as Latin and Ancient Greek, in which exemplary paradigms are rote-learned and principal parts are provided alongside new vocabulary items. The implicational relations between the principal parts provide a language-user with sufficient information to determine the inflexion class to which a novel lexeme belongs, and knowledge of the full paradigm of

[14] Ackerman et al. (2009).

just one exemplary lexical item permits, theoretically, the correct deduction of any unknown form in the novel lexeme's paradigm.[15]

As has been demonstrated by Ackerman and Malouf (2013), cross-linguistically even those languages characterized by a great deal of surface complexity (the 'Enumerative' or 'E-complexity', referring to "the number of morphosyntactic distinctions that a language makes and the strategies employed to encode them, concerning either the internal composition of words or the arrangement of classes of words into inflectional classes") tend to exhibit low conditional entropy on average.[16] That is, reference to at least one other fully-inflected wordform in the inflexional paradigm of a lexeme frequently reduces uncertainty, often quite drastically, about another form in the paradigm. This is referred to as the 'Low Conditional Entropy Conjecture', "the prediction that all languages will show low average conditional entropy",[17] which provides a partial explanation for why seemingly very complicated systems can be learned and used by L1 speakers by with just as much ease as L1 learners of languages with low E-complexity.

Although entropy is a useful notion, it is relatively new and there are still some problems with the ways in which entropy measures are computed. In particular, as pointed out by Bonami and Beniamine (2016), the calculation of the conditional entropy requires the segmentation of stems and exponents which, as will be discussed further in this chapter, is often arbitrary. The choices made in the splitting of lexical stem and inflexional exponent can drastically affect the average entropy of the morphological case paradigm. Furthermore, since entropy calculates the amount of uncertainty associated with variation in inflexion, in its

[15] It is unlikely that native speakers of a language 'store' whole exemplary paradigms (except perhaps for lexemes which are particularly frequent). Instead, a native speaker may recognize the implicational relations that exist between two forms for a first declension lexeme {PUELLA} (e.g., PUELLA ~ PUELLĀS), a different set of abstract patterns for another first declension lexeme {MENSA} (e.g., MENSA ~ MENSAE ~ MENSĪS) and so on. The sum of these provide sufficient information for a native speaker to deduce all forms of a nominal belonging to this inflexion class, though not necessarily in quite the same way as the L2 learner.
[16] Ackerman and Malouf (2013:429).
[17] Ackerman and Malouf (2015:6).

current guise it cannot capture the concomitant uncertainty and certainty associated with sub-word formal variation in a language such as Finnish. That is, entropy measures can calculate the amount of uncertainty associated with the different alternations exhibited by the nominal stem (§3.1.1); it can also calculate, separately, that the genitive singular exponent is a minimally-entropic suffix *-n*. But the current entropy equations do not accommodate the predictiveness afforded by minimally-entropic sub-word material, as they are sensitive only to variance *or* invariance, not a mixture of the two.

One final aspect of modern WP that must be considered briefly here is that of 'irregular' versus 'regular' morphology.[18] In constructivist approaches, regularity is considered the 'norm', as it provides the speaker and the hearer with information about the production and interpretation respectively of novel inflected forms. Irregular forms need to be individually accounted for by postulating either form storage or (usually *ad hoc*) (morpho)phonological processes which affect an 'underlying' form. In an abstractivist approach, on the other hand, it is regularity, not irregularity, which requires explanation. In the first place, highly irregular - such as suppletive - forms are actually easier for hearers to discriminate from other forms in the speech stream and, therefore, associate with distinct functions. Furthermore, various psychoacoustic tests have revealed that there are durations differences between apparently homophonous forms in English, including below (Plag et al. 2016) and at the level (Gahl 2008 and Drager 2011) of the word. Further tests may reveal that those recurrent partials which appear to be the 'same', and which are therefore described as 'morphemes' in IA and IP, for instance, are actually distinct. Moreover, research into sub-phonemic durational distinctions between inflected items indicates that "acoustic differences exist between uninflected and inflected forms and that listeners are sensitive to them".[19] Even when they are

[18] Discussed in detail by Blevins et al. (2017).
[19] Kemps et al. (2005:441).

'regular', the production of novel forms involves more than just the simple concatenation of lexical and inflexional material.

'Irregularity' in an abstractivist model is often considered to play a 'discriminative' role. Discriminative models of language learning and language use, such as those developed by Ramscar and colleagues (e.g., Ramscar and Dye 2010, Ramscar et al. 2010 and Ramscar 2013), argue that "the main function of phonology is to discriminate between semantic alternatives".[20] Thus, where a language-user encounters inflexional variation within a paradigm (or, indeed, anywhere), they associate it with some difference in function.[21] Even where forms are not consistently distinguished by the same sub-word exponents - for instance, where 'genitive plural' functions are differentiated from other forms in the paradigm by means of a suffixal marker *-ārum* in one declension and *-um* in another (as in Latin) - it does not matter that these sub-word elements are phonologically distinct. Instead, it matters that the functional relationship between PUELLĀRUM and PUELLAE is the same as that between MILITUM and MILITĒS. In this way, distinct sub-word units may discriminate the same semantic (stylistic, regional etc.) differences within the language.

Furthermore, "[e]lements that serve principally to discriminate larger forms need not have a single function or meaning in all of the contexts in which they occur, but may perform different discriminative functions in different contexts."[22] In Livonian, for instance, the suffix *-õ* 'encodes' different functions depending on the shape of other inflected forms in the paradigm. In examples such as *sieldõ* 'clear', *aššõ* 'sudden, abrupt' and *drūʼošõ* 'brave', *-õ* is found in the nominative/genitive/partitive singular forms, which contrast with other inflected items distinguished by additional suffixal markers signalling a difference in function, e.g.,

[20] Ramscar and Dye (2010:28).
[21] In this way, the 'Principle of Contrast' ("Every two forms contrast in meaning" - Clark 1987:2), Synonymy Avoidance, and 'No-Blur'/Vocabular Clarity (Carstairs-McCarthy 1994, 2010) may all be considered to fall under the purview of discriminative language learning.
[22] Blevins (2016:211).

aššõ-n - dative singular of {AŠŠÕ}; *drū'ošõ-d-õks* - translative-comitative plural of {DRŪ'OŠÕ}. By contrast, in nominals such as (nominative singular) *rikāz* 'rich' and *āmbaz* 'tooth', the final *-õ*, along with an alternation in the stem shape, serves to discriminate the genitive singular from the nominative singular form, i.e., genitive singular *rikkõ* and *ambõ*. Finally, the partitive/illative singular form *ka'llõ* 'fish' is distinguished from other forms in the paradigm by means of a stem change in addition to the suffix *-õ* (cf. nominative/genitive singular form *kalā* and the partitive plural form *ka'ļḍi*). In this way, it is not a desinence *-õ* itself, but the relationship between a form ending in *-õ* and other forms in the lexeme's paradigm, which distinguishes particular functions.

Consideration of the pressures of *irregularity* and *regularity* is essential to understanding the solving of the Paradigm Cell-Filling Problem. Regularity aids in the prediction of novel forms, whilst irregular forms can impede the deduction of correct inflected forms as they are not easy to predict given knowledge of (certain) other forms in the paradigm, but can be informative about the functions a form encodes. The competing pressures of regularity and irregularity thus provide a language-user with predictive and discriminative information respectively.

It should be noted here that languages with sub-word forms which are minimally-entropic – particularly in Finnish and Veps - provide a language-user with a great deal of *predictiveness*. Case and number affixes in these varieties can be identified, abstracted away from fully-inflected wordforms, associated with particular functions, and stored; whilst they serve to discriminate inflected forms in the inflexional paradigm, they can also provide certainty in the inflexion of unknown forms.

Despite the apparent strengths of the modern WP perspective, it will later be suggested that, in their current guise, modern abstractivist accounts fail to recognize important

qualitative differences between minimally-entropic and non-minimally-entropic sub-word formal variation. Consideration of these differences shed light on the reason why some languages maintain or even propagate whilst others lose morphological case distinctions.

4.1.3 CLOSING REMARKS

There are many instances in which the direct mapping of function and form is disturbed in the languages considered so far (as will be discussed in more detail in the sections following), and these pose problems for morphemic accounts in particular. When words are conceived of as being built up from sub-word elements by means of concatenation of morphs or by rules licensing the introduction of particular units of exponence, often something is lost. Most frequently in natural language, it transpires that an inflected wordform is 'greater than the sum of its parts', and that it is only with a network view of the ways in which inflected items 'relate' to other forms in the language that we can understand how a case system develops.

In §§4.2-3, I will describe how the data from the Finnic and Italic languages, particularly those aspects which prove problematic for constructivist approaches, may be easily accommodated by WP. For reasons of space, I will not offer an exhaustive constructivist analysis of each of the varieties discussed in this section. Instead, representative samples of the sorts of inflexional patterns found in the Italic and Finnic languages will be considered, the conclusions from which may be extended to other parts of the case systems of these varieties.

4.2.0 THE ITALIC LANGUAGES

In this section I will examine the data presented in chapter two and offer an analysis of the ways in which the inflexional structures of the Latin, old French, old Occitan, and Romanian case systems may be accommodated by constructivist and abstractivist perspectives. I will focus here on instances in which the form~function mapping is disturbed in these varieties -

including problems inherent in the segmentation of inflected forms, formal variation in the stem and in the suffix across number or declension (for instance), and syncretism - in order to demonstrate the weaknesses and strengths of the different models.[23]

4.2.1 '(DE-)COMPOSITIONALITY' AND ATTENDANT ISSUES

Constructivist accounts frequently encounter problems when it comes to identifying invariant sub-word elements (whether this is the lexical stem or inflexional material) which 'encode' particular functions. Numerous challenges are posed by fusional morphology, in which the form~function mapping is opacified in various ways. When it comes to decomposing these forms, there are frequently issues in which the identification of morphemes is often arbitrary and, sometimes, impossible.

In the first place, in a small sub-section of the Latin case system - the singular forms of first declension nominals - decomposing forms into 'lexical stem' (bearing lexical meaning) and 'inflexional suffix(es)' (bearing case, number, and gender functions) may initially seem straightforward.[24] A relatively uncontroversial analysis of the forms of the lexeme {PUELLA} would recognize a stem shape *puell-*, and analyse the phonological material which follows as the inflexional case, number, and gender markers:

[23] Some of these Italic data have already been considered from one or several constructivist perspectives. Kihm (2012), for instance, has discussed the old French and Romanian case systems from the Paradigm Function Morphology perspective, and Matthews (1972) offers a very comprehensive argument against the morpheme-based models of morphological description for Latin verbs.

[24] Although further division of inflexional suffixes into 'case' and 'number' (and 'gender') morphs is frequently impossible. This is a problem which will be discussed later in this section.

Singular	Lexical stem	Inflexional desinence
Nominative	*puell*	*a*
Accusative	*puell*	*am*
Genitive	*puell*	*ae*
Dative	*puell*	*ae*
Ablative	*puell*	*ā*

Table 4.4: Decomposing the singular forms of Latin first declension {PUELLA}[25]

What is more problematic is the further decomposition of the inflexional material in the rightmost column of table 4.4 into separate markers of 'case' and 'number' values.[26]

	Stem	Number	Case
Nominative	*puell*	*a*	*Ø*
Accusative	*puell*	*a*	*m*
Genitive	*puell*	*a*	*e*
Dative	*puell*	*a*	*e*
Ablative	*puell*	*a*	*a/-V/L*

Table 4.5: Decomposing the singular forms of {PUELLA}[27]

Here, the element *-a-* is common to all forms and might therefore be analysed as the 'singular' marker, but the case suffixes identified present problems for a morphemic analysis. In the nominative singular, for instance, there is no overt phonological form besides the final *-a* (which, according to the analysis above, is associated with singular number), which proves challenging for model such as IA in which every function must be expressed by an overt phonological form.

Additionally, the ablative singular form is characterized by a long final vowel. In IA, we could suggest that the form of the ablative case is *-a* and that this, in combination with the

[25] The following analyses are based mainly on orthography. This reflects a further weakness of many constructivist accounts, which often consider written rather than spoken data.
[26] In the paradigms of adjectives, gender is also cumulatively expressed by suffixal forms. This simply introduces further problems of analysis for a morphemic approach to the data.
[27] This analysis does not pretend to be uncontentious. The fact that there is more than one way to split the inflexional suffixes of first declension nominals and that each may be equally well-motivated simply highlights the frequently arbitrary nature of decomposition, a major weakness inherent not just in morphemic accounts, but in constructivist approaches more generally.

singular morph *-a-*, gives *-ā*.[28] This analysis, however, does not work outside of the first declension and so, instead, it might be argued that the ablative morph is *-V*, a copy of the vowel sound immediately following, or that this form is derived, in IP, by means of a lengthening process (the '*L*' in table 4.5).[29] Since ablative singular forms across the language are frequently characterized by a form ending in a long vowel, both the IP analysis of the vowel-lengthening process and the identification of a morph *-V* may have some generality across the language. However, the fact that there are two equally well-motivated hypotheses, and no principled means of deciding between them, introduces a further limitation of constructivist approaches.

A morphemic analysis meets further problems when the plural forms of {PUELLA} are considered:

	Stem	**Number**	**Case**
Nominative	*puell*	*a*	*e*
Accusative	*puell*	*ā*	*s*
Genitive	*puell*	*ā*	*rum*
Dative	*puell*	*ī*	*s*
Ablative	*puell*	*ī*	*s*

Table 4.6: Analysing the plural forms of {PUELLA}

Here, three 'number' exponents are identified: *-a-*, *-ā-*, and *-ī-*, which violates the morphemic ideal and so, in order to save an IA or IP analysis, an 'underlying' plural morph could be posited, the surface form of which changes depending either on a) grammatical or b) phonological context. The former is problematic. Either we need to identify a 'natural class' of features which 'select' the different allomorphs (for instance, [+oblique]) - something

[28] In this analysis, the ablative marker is therefore the same as the singular number exponent, but homonymy is perfectly permissible in a morphemic account, as there is a requirement that form and function be in an isomorphic mapping, but not necessarily that every function be encoded by a phonologically distinct form.
[29] This latter element is "quite unlike any genuine phonological element" and is more like a "notational device which will 'trigger off'" the rule of lengthening the preceding vowel. Matthews (1972:63).

which is difficult for the distribution of the plural morphs *-ā-* and *-ī-* - or simply enumerate the grammatical environments in which each allomorph occurs, an inefficient alternative.

To sidestep the difficulties associated with the identification of 'natural classes', and with simply listing the seemingly random array of case values which 'select' particular plural allomorphs, we may instead suggest that there is a phonological environment which causes the underlying plural morpheme to change. For instance, *-a-* occurs before vowels (as in the nominative plural form), *-ī-* before *-s* (as in the dative and ablative plural forms), and *-ā-* elsewhere (as in the genitive plural form). However, this analysis does not work for the accusative plural form, where *-ā-* is followed by *-s*, just like the form *-ī-*. We might save our analyses by suggesting instead that the accusative plural case-form is *-Ls* rather than *-as* (where '*L*' stands for a similar lengthening process as was discussed in reference to the ablative singular form). In this instance, a morphemic analysis might still work, since the '*L*' constitutes an 'elsewhere' environment licensing the selection of the allomorph *-a-* (and subsequently causing its lengthening). However, if we posit such a solution, it is also possible to suggest that the genitive plural form is *-Lrum* rather than *-rum* and that *-a-* is the underlying plural morph in the nominative, accusative, and genitive forms. This improves generality, but causes problems: what is to stop us from suggesting that the morpheme alternant of the plural is *-i-* rather than *-ī-*, and that the dative plural/ablative plural suffix is *-Ls*? How do we determine which is the 'correct' hypothesis in a principled way?

Across number, in an IA account, feature values should be expressed by markers which are the same throughout the language (unless altered regularly by the phonological environment). In the Latin first declension,[30] however, we find no such structure:

[30] And, indeed, in the other major declensions.

	Singular case-forms	Plural case-forms
Nominative	*Ø*	*e*
Accusative	*m*	*(L)s*
Genitive	*e*	*(L)rum*
Dative	*e*	(*L*)*s*
Ablative	*a/L*	(*L*)*s*

Table 4.7: Comparing first declension case morphs across singular and plural

Across number, according to the analyses offered here, there are no instances in which the case markers are identical. Once again, it might be suggested that the distribution of these morpheme alternants is phonologically or grammatically conditioned. But the phonological explanation falls down upon consideration of the distribution of the nominative case markers *-Ø* and *-e*, which are found in the same phonological environment, immediately following the stem *puell-* and the (number?) marker *-a-*. If it is argued that the grammatical environment of [+plural], for instance, 'selects' the nominative allomorph *-e*, the accusative allomorph *-(L)s*, etc., we are required to list allomorphs for each of the case values in different inflexion classes as well as the environments in which they occur. This offers no generality across the language and nullifies what Bloomfield (1933:238) refers to as the 'scientific compactness' of an atomistic model.

Another problem inherent in the decomposing of forms into sub-word parts is the frequently arbitrary splitting of lexical stem and inflexional affix. In the Latin example of first declension singular nominals, it was concluded that the stem shape must be *puell-* and the rest of the form the 'inflexional material' (regardless of whether or not these forms could be further decomposed), since this stem shape is found throughout the paradigm. However, it is not always so simple to identify uncontentious 'stem' versus 'inflexional' material. This is demonstrated by the following Romanian examples:

	{ZI} 'day'		{STEA} 'star'	
	Singular	**Plural**	**Singular**	**Plural**
Nom.-acc.	*zi*	*zile*	*stea*	*stele*
Gen.-dat.	*zile*	*zile*	*stele*	*stele*

Table 4.8: Morphological alternations in the paradigms of Romanian {ZI} and {STEA}

There are two ways in which these forms are generally decomposed into 'lexical stem' and 'inflexional material' in Romanian grammars: *zi-Ø ~ zi-le* (*stea-Ø ~ ste-le*),[31] and *zi-Ø ~ zil-e* (*stea-Ø ~ stel-e*).[32] The first of these alternatives requires us to analyse a form *-le* as inflexional material, even though, when compared to other suffixal markers found in the nominals of the language, it is very unusual (other suffixes in feminine nominals usually consist solely of vowels or begin with vowels). The second of these analyses necessitates the recognition of a stem alternation between the nominative-accusative singular cell and the form found throughout the rest of the paradigm. There are supporting arguments for both of these analyses, and neither can be considered 'wrong' or 'right', meaning that there is no principled way of splitting these forms up, even on consideration of the inflexional behaviour of other lexemes in the language. Without a principled splitting of fully-inflected wordforms, how is it possible to associate sub-word parts with particular functions?

Unlike the morphemic models IA and IP, an abstractivist account does not necessitate the identification of stems and exponents or of sub-word units which invariably encode particular functions. Instead, the fully-inflected wordform is analysed as 'basic', and need not be further decomposed into isolable and meaningful sub-word elements. For instance, in the example of {PUELLA} discussed above, it is not required that plural number be expressed by phonologically or grammatically conditioned allomorphs of an underlying form which are somehow 'derived' depending on the environment in which they occur. Indeed, no sub-word

[31] This segmentation is preferred by synchronic analyses of the Romanian language such as Guțu Romalo (1968:57), Nedelcu (2013b:275).
[32] This segmentation is preferred by Maiden (2016e).

units need be identified at all. Rather, the fully-inflected wordform, the relationship it bears to other forms in the paradigm, and the functions it encodes provide language-users with information about the other forms in the inflexional paradigm. Therefore, in Latin, the form PUELLA, not the combination of a stem *puell-* and an inflexional suffix *-a*, is analysed as encoding 'nominative singular' functions because it stands in a certain implicative relation to other forms in the paradigm.

In Romanian, similarly, it is the abstract pattern existing between the forms *zi* and *zile* (in addition to the functions they encode), not whether they are split up as *zi ~ zi-le* or *zi ~ zil-e*, that provides information about other forms in the language. For instance, on the model of the implicative patterns exhibited by *stea* and *stele*, we may derive the feminine genitive-dative singular form of the adjective {GREA} 'hard', without necessitating the arbitrary splitting of 'lexical stem' and 'inflexional affix':[33]

stea : stele

grea : ?

? : grele

Figure 4.5: The analogical development of Romanian genitive-dative singular *grele* 'hard'

Both of the nominative-accusative singular forms *stea* and *grea* end in the rising diphthong *-e̯a*; all other nominative-singular forms in the language end in this diphthong are characterized by a genitive-dative singular form (and, indeed, a nominative-accusative/ genitive-dative plural form) in *-le*. On the model of alternations exhibited by these nominals in the same lexical neighbourhood (i.e., those which end in *-e̯a* in the nominative-accusative

[33] This is what has happened historically, since the *-l-* in the form *grele* is unexpected on consideration of the Latin lexeme {GRAUIS}, the forms of which contained no sound which could regularly have become /l/ in Romanian. The Romanian adjective {GREA} has joined the inflexion class of feminines ending in a stressed vowel by analogy with other forms with a similar structure. cf. also *mea* 'my' (feminine), pl. *mele*. My thanks to Martin Maiden (p.c.) for bringing this to my attention.

singular), the unknown second form of the adjective {GREA} may be deduced. There is no need to split this form up nor to identify sub-word units which 'encode' particular functions and which alter according to phonological or grammatical environments. Instead, in an abstractivist account, the whole wordform and the relationship it bears to other forms in the inflexional paradigm (as well as in the wider lexical neighbourhood and the morphological family) provide sufficient information for language-users to deduce previously unencountered forms.

4.2.2 ALLOMORPHY

The term 'allomorph' has, prior to this chapter, been reserved for formal variants of the lexical stem alone; phonologically distinct exponents which can be associated with the same case value across number and/or declension have not been described as 'allomorphy in the case suffix'. The term 'allomorph' is inherently referential: a form can only be considered an 'allomorph' if another form can be associated with the same feature values. In constructivist approaches such as IA and IP, where underlying morphs are posited, the term 'allomorph' may safely apply both to formal variation in the stem and in inflexional material. According to an approach (such as WP) that recognizes inflexion classes as central organizing structures of a language (rather than epiphenomena), however, it is not strictly accurate to describe, e.g., the Latin suffixes *-ae, -ī, -is, -ūs, -(e)ī* as 'allomorphs' of the genitive (singular) case. These different exponents, although they can all be associated with 'genitive' and 'singular' values, differ in respect of the inflexion class in which they occur; in Latin, there is only one *first declension* genitive singular exponent (the form *-ae* is not found alongside *-ī*, for instance, in the same lexemes). For this reason, it is only in reference to constructivist accounts that I will

use the terms ‘allomorph’ and ‘allomorphy’ to discuss case exponents.[34] Since formal variation in the stem is found *within* the paradigm of a lexeme, I will continue to use the terms ‘allomorph’ and ‘allomorphy’ in reference to lexical stem shapes in both the Italic and Finnic languages.

As has been extensively noted, formal variation, whether in the stem (within the declension) or in the suffix (e.g., across number or declensions), presents a clear instance in which the form~function mapping is interrupted. That is, there are multiple sometimes very distinct shapes encoding the ‘same’ case/number functions. Each of the Italic languages introduced in chapter two exhibits allomorphy in the lexical stem. In old French and old Occitan, one inflexion class for masculine and one for feminine lexemes includes all imparisyllabic nominals:

	Old French	
Masculine *ber* - baron	**Singular**	**Plural**
Nominative	*ber*	*baron*
Oblique	*baron*	*barons*
Feminine *suer* - sister	**Singular**	**Plural**
Nominative	*suer*	*serors*
Oblique	*seror*	
	Old Occitan	
Masculine *trobaire* ‘troubadour’	**Singular**	**Plural**
Nominative	*trobaire*	*trobadór*
Oblique	*trobadór*	*trobadórs*
Feminine *sór* - sister	**Singular**	**Plural**
Nominative	*sór*	*seròrs*
Oblique	*seròr*	

Table 4.9: Imparisyllabic forms in old French and old Occitan nouns

In these lexemes, there is a change in the nature of the vowel in the stem and an addition or loss of phonological material that cannot necessarily be described in a straight-forward

[34] In reference to certain suffixal elements in particularly Finnish, the term ‘allomorphy’ could be used of certain suffixal shapes, where, for instance, the exponents *-A* and *-tA* can, in the inflexional paradigms of trisyllabic nominals ending in *-eA*, both be used to discriminate ‘partitive singular’ functions.

way as a ‘case’ or ‘number’ desinence, since it is not consistent across different nominals. There are two ways in which a morphemic approach may accommodate this sort of stem allomorphy. In IA, two stem shapes may be lexically stipulated, along with the environments in which they are ‘selected’. It might therefore be suggested that speakers of old Occitan know and store two stem shapes *sór* and *seròr*, and that the former is selected in the grammatical environment ‘nominative singular’ and the latter occurs elsewhere.

Alternatively, in an IP account, it could be argued that the allomorphy in the stem is the result of a morphological ‘process’ affecting an underling form. For example, we might suggest that *baron* is the underlying form of the lexeme {BER} (as it is found in a majority of the inflexional cells in the paradigm), and that the nominative singular form *ber* is produced by means of phonological processes when the nominative morph is introduced in the singular. Here, therefore, the surface form *ber* represents that ‘nominative’ and ‘singular’ process morphs have affected the underlying stem of the lexeme {BER}. However, it is not clear why the addition of these process morphs licences this change in the nominative singular and not in other forms characterized by ‘nominative’ or ‘singular’ process morphs.

Where allomorphy in the stem is found in an inflexional paradigm it must be accounted for either by positing phonological rules to account for different surface forms or by arguing that different stem shapes are ‘listed’ in the lexicon. As will be discussed later in the section, however, where a model does not necessitate the splitting of ‘lexical stem’ and ‘inflexional material’, allomorphy may be considered to play a discriminative role in language use, enabling greater communicative efficiency, whilst the distribution of distinct forms within the paradigm often provides information which may be used in the production of novel forms.

In Latin, the nominative singular form of third declension (non-neuter) nominal is sometimes characterized by a different stem shape to that found in all other forms in the paradigm:

	{RĒX} 'king'		**{URBS} 'city'**		**{NOX} 'night'**	
	Singular	**Plural**	**Singular**	**Plural**	**Singular**	**Plural**
Nominative	RĒX	RĒG-ĒS	URBS	URB-ĒS	NOX	NOCT-ĒS
Accusative	RĒG-EM	RĒG-ĒS	URB-EM	URB-ĒS	NOCT-EM	NOCT-ĒS
Genitive	RĒG-IS	RĒG-UM	URB-IS	URB-IUM	NOCT-IS	NOCT-IUM
Dative	RĒG-Ī	RĒG-IBUS	URB-Ī	URB-IBUS	NOCT-Ī	NOCT-IBUS
Ablative	RĒG-E	RĒG-IBUS	URB-E	URB-IBUS	NOCT-E	NOCT-IBUS

Table 4.10: Stem allomorphy in Latin {RĒX}, {URBS}, and {NOX}

It might be possible to give a phonological account of this stem allomorphy. For example, although it is not clear from the spelling of the lexemes {RĒX} and {NOX}, the final phoneme of the nominative singular form of all of the lexemes in table 4.10 is *-s*. We may analyse this as the 'nominative' morph which is concatenated to the stem shapes *rēg-*, *urb-*, and *noct-*, the stem shapes found in the rest of the respective paradigms. For RĒX and URBS-type nominals, the [-voiced] sibilant *-s* leads (via regular phonological change) to the de-voicing of the stem-final consonant, where [+voiced] ⟶ [-voiced] / ____ *s*.[35] For the lexeme {NOX}, we find a slightly different phonological change C ⟶ Ø / C ___ *s*.[36] These phonological processes thus accommodate for the stem allomorphy exhibited by these lexemes. However, these rules do not work for all lexemes in which allomorphy in the stem is found:

{SUPELLEX} 'furnishings, furniture'	**Singular**	**Plural**
Nominative	SUPELLEX	SUPELLECTILĒS
Accusative	SUPELLECTILEM	SUPELLECTILĒS
Genitive	SUPELLECTILIS	SUPELLECTILUM
Dative	SUPELLECTILĪ	SUPELLECTILIBUS
Ablative	SUPELLECTILE	SUPELLECTILIBUS

Table 4.11: Stem allomorphy in Latin {SUPELLEX}

[35] Sihler (1995:203f.).

[36] These 'solutions' certainly do not work generally across the language. In addition to the problem posed by {SUPELLEX} below, we find examples such as HOMŌ ~ HOMIN- and HOSPES ~ HOSPIT-, which must be accounted for by means of different phonological rules.

As in the lexemes discussed above, the nominative singular form of {SUPELLEX} ends in *-s*. If the 'second' stem were **supelect-* or **supeleg-*, on the model of {NOX} and {RĒX} respectively, one of the rules posited above would be sufficient general to 'derive' one stem shape from another. However, the stem form found throughout the paradigm is actually *supelectil-*. How do we account for the formal variation exhibited by this lexeme?

One solution might be to list, rather than derive, the two stem allomorphs for {SUPELLEX}: *supellect-* for the nominative singular and *supellectil-* elsewhere.[37] These different stems and their selection must be stipulated in the lexicon, since they cannot be derived by referring to the regular phonological processes posited for the nominals in table 4.10. Alternatively, on a phonological account, we might suggest that the stem allomorph *supellect-* is selected before final *-s* and elsewhere the stem is *supellectil-*. Problematically, is not clear where we should draw the line between grammatically-selected and phonologically-conditioned allomorphy, nor how to determine in a non-arbitrary way which analysis is 'right' for a given form. In instances of suppletive stem allomorphy, different stem forms need to be listed. In the example given here in reference to Latin, however, the two stem forms *supellect-* and *supellectil-* are more similar than they are distinct (they differ only in the final *-il-* of the second stem). Surely these forms are formally related enough for one to be 'derived' from the other (or from a distinct underlying form)? But, since there are not, to the best of my knowledge, any other lexemes in Latin which exhibit this type of alternation, any process offered here cannot be described as a general or productive one. Instead, either these stem allomorphs must be listed despite their overwhelming phonological similarity, or *ad hoc* phonological processes which do not appear to have any generality in the modern stage of the language must be posited.

[37] The nominative singular stem is analysed as *supellect-* here as this form is found in the second stem *supellectil-*.

Of the Italic languages discussed here, Latin exhibits the most extensive formal variation in the inflexional suffix and, within the Latin declensions, it is impossible to identify invariant suffixes which encode particular case values across number; across declensions, it is rare to find the same inflexional exponents encoding the same case(/number) values:

	First Declension		**Second Declension**		**Third Declension**	
	Singular	**Plural**	**Singular**	**Plural**	**Singular**	**Plural**
Nom.	PUELLA	PUELLAE	SERUUS	SERUĪ	RĒX	RĒGĒS
Acc.	PUELLAM	PUELLĀS	SERUUM	SERUŌS	RĒGEM	RĒGĒS
Gen.	PUELLAE	PUELLĀRUM	SERUĪ	SERUŌRUM	RĒGIS	RĒGUM
Dat.	PUELLAE	PUELLĪS	SERUŌ	SERUĪS	RĒGI	RĒGIBUS
Abl.	PUELLĀ	PUELLĪS	SERUŌ	SERUĪS	RĒGE	RĒGIBUS

Table 4.12: Formal variation of the suffix across the first, second, and third declensions in Latin[38]

The form *-s* is found in the {RĒX} nominative singular and nominative plural forms, as well as the nominative singular form of {SERUUS}, but it cannot be analysed as a marker of the 'nominative' in a straight-forward way, as it does not occur in the nominative forms in either number in the first declension, nor in the plural form in the second declension. It might be suggested instead that the 'nominative' morph is something other than *-s* (although finding another sub-word form that is overt and consistent across both number *and* declension class is impossible, at least in the nominative). Alternatively, it might be argued that the nominative *-s* is somehow actually underlyingly present in the nominative forms of the first declension and the nominative plural of the second declension, but that there are various processes that generate surface forms without this overt morph *-s*. For instance, we might suggest that the nominative morph *-s* is deleted regularly following diphthongs and short vowels, although not following consonants or long vowels. This analysis works for the first declension

[38] The following discussion is restricted to these three declensions and does not take into account, e.g., neuter forms in the second and third declensions or other third declension nominals which do not end in *-s* in the nominative singular; the fourth and fifth declensions are also omitted from consideration. Introducing these different forms would simply offer yet more evidence against a morphemic account of the Latin case system.

nominative singular and plural forms as well as the forms in the third declension. However, the form SERUĪ proves problematic, since *-ī* is not the environment in which we have suggested that final *-s* to be lost. Accounting for the extensive suffixal allomorphy of the Latin case system from a constructivist perspective, then, is likely to lead to some very strange 'solutions', many of which may in fact violate the spirit of a model that prioritizes direct sub-word form~function mapping.

In an abstractivist account, the implicational informativeness associated with the formal variation in the suffix in Latin nominals is highlighted. Frequently, the abstract relations between suffixal alternations reduce uncertainty about the way in which a lexeme inflects, as well as, sometimes, even the gender of a nominal. In fact, when it is not required that suffixal marking be in a one-to-one form~function mapping, the Latin case system is revealed as one "où tout se tient",[39] where each form (in addition to the functions it encodes) often provides (near) minimally-entropic information about the remaining forms in the paradigm. For instance, where an inflected form ends in *-a*, such as BELLA 'wars' (encoding nominative/accusative plural functions), it provides minimally entropic information about the gender of the lexeme (it can be nothing other than neuter), even though, heretofore, this suffixal morph has only been discussed in reference to the nominative singular forms of the first declension nominals. On the other hand, where a language-user meets a form BELLA which encodes nominative singular functions, they may correctly deduce that it does not belong to the paradigm of the lexeme {BELLUM}, but instead is the feminine nominative singular form of the adjective {BELLUS} 'pretty', since a form ending in *-a* and encoding nominative singular functions cannot be neuter. It does not matter that, whatever way we deconstruct these forms, the inflexional material *-a* will end up being associated with

[39] Meillet (1903:407).

different functions: in WP, it is important what relation the fully-inflected wordform bears to other forms in its paradigm, not what the 'sum' of its sub-word parts is.

4.2.3 SYNCRETISM

Syncretism is yet another way in which the form~function mapping can be interrupted in language: where two forms are expected, only one occurs encoding two (sets of) functions. One striking instance of syncretism is that found in the Romanian inflexional case system, where the genitive-dative singular form is almost invariably the same as the form found for both cases of the plural.

	Singular	**Plural**
Nominative-accusative	FORM1	FORM2
Genitive-dative	FORM2	FORM2

Table 4.13: Syncretism in (most) Romanian feminine nouns

Even if there were some way to 'construct' the genitive-dative form from the root of the nominal (if, indeed, it is actually possible to non-arbitrarily separate 'lexical stem' and 'inflexional material' in Romanian), we have to in some way account for the fact that the same form encodes genitive-dative singular, nominative-accusative plural, *and* genitive-dative plural functions.

In an IA account, the differences in function may be analysed as being expressed by zero-morphs, though these are inherently problematic for those models which prioritize *forms* which are in an isomorphic relationship with function. It may instead be suggested that the forms are distributed according to a 'natural class' of feature values. That is, that case is neutralized in the grammatical environment of [+plural] (or, alternatively, that number is neutralized in the grammatical environment of [-nominative-accusative]). However, the combination of the distinct functions which are encoded by [FORM2] in table 4.13 means that neither of these natural class analyses work.

The common pattern of syncretism in Romanian feminine nominals, therefore, is problematic for a strict IA approach. IP might accommodate this inflexional pattern by positing that the genitive-dative singular form may derived by applying an operation on an underlying form of a nominal which generates the relevant surface form. For instance, the nominative genitive-dative singular of the lexeme {FATĂ} may be derived by applying an operation to the underlying stem which generates the surface form *fete* associated with genitive-dative singular functions. A different process may then be posited which generate the same form *fete* which is associated with nominative-accusative plural functions, and yet another which would generate the same form with genitive-dative plural functions.[40] Similar issues attendant on the introduction of zero process morphs as were discussed in reference to the derivation of old French and old Occitan stem forms (§4.2.2) may be identified here, however. Why does the introduction of the nominative process morph not result in the shape *fete* in the singular? And why does the introduction of singular process morph in the nominative result in the form *fată* as it does in the genitive-dative singular?

Syntagmatic approaches obscure the often informative nature of systematic patterns of syncretism within the inflexional paradigm. In fact, when considered as whole wordforms, and not decomposed into lexical stems and inflexional material, these patterns can reduce uncertainty about the shapes of unknown inflected forms. According to WP, the morphomic distribution of forms exhibited by feminine forms in Romanian might be considered informative. After all, the distribution of these two forms is very regular throughout the language; if a speaker knows the form of the genitive-dative singular, they also know, almost for certain, that of the plural, too (and vice versa).

[40] Alternatively, we could suggest that *fete* is the 'underlying' form, since it occupies the most cells in the inflexional paradigm, and that the nominative-accusative singular form *fată* is the one derived. In this instance, the nominative-accusative plural form would be 'generated' by a process morph that applies no operation to the base form and maps it to genitive-dative plural functions etc.. However, this analysis highlights the arbitrary nature of selecting 'underlying' and 'derived' forms.

The importance of the paradigmatic distribution of fully-inflected wordforms in determining the functions they encode can be seen particularly by comparing Romanian with old French and old Occitan.

	Masculine				**Feminine**			
	***murs* 'wall'**		***paire* 'father'**		***maire* 'mother'**		***nau* 'ship'**	
	Singular	**Plural**	**Singular**	**Plural**	**Singular**	**Plural**	**Singular**	**Plural**
Nominative	*murs*	*mur*	*paire(s)*	*paire*	*maire*	*maires*	*nau(s)*	*naus*
Oblique	*mur*	*murs*	*paire*	*paires*	*maire*	*maires*	*nau*	*naus*

Table 4.14: Selected old Occitan nominal paradigms

According to the grammars consulted, masculine nominals in both old French and old Occitan more consistently distinguish case in both the singular and the plural than feminine nominals. In old Occitan, for instance, in the largest class of feminine nominals (*maire*), case is not distinguished in either number, whereas in the largest class of masculine nominals, case is consistently marked by the alternation of *-s* (or *-z*) and *-Ø*. The cumulative nature of the suffix *-s/-z* and the lack of overt marking with which it alternates are both problematic in a morphemic approach, but reference to the paradigmatic distribution of whole wordforms ending in *-s/-z* is more informative about the deduction of novel forms than syntagmatic 'parts'. For instance, the paradigmatic distribution of wordforms in the paradigm is the same for all lexemes belonging to the *murs* class in both old French and old Occitan:[41]

	Singular	**Plural**
Nominative	FORM1	FORM2
Oblique	FORM2	FORM1

Table 4.15: Chiastic distribution of *murs*-type nominal forms (old French and old Occitan)

In order to know the entire paradigm for *murs*, therefore, a language-user need only know two forms, and distribute them according to the pattern given in table 4.15. This helps to increase the predictability of the shape of previously unencountered forms.

[41] Although it will be repeated that inflexion class membership in old French and old Occitan was not quite as 'fixed' as it appears to have been in particularly Classical Latin.

However, there is a 'weakness' in the predictability of the inflexional patterns in old French and old Occitan nominals, as the 'same' suffix *-s/-z* is found in all other paradigms, albeit distributed in different ways. In reality, in order to determine the inflexion class of a lexeme, as many as three forms and their functions need to be known. For instance, where an old French language-user encounters a nominative singular form *voisins* 'neighbour', without reference to another form of this lexeme, there is not sufficient information to determine whether this nominal follows the inflexional pattern of *murs*-type or imparisyllabic nominals (whether, for instance, the oblique singular form is *voisin* or perhaps a completely different stem shape). Similarly, from knowledge of the oblique singular form *voisin* alone, a speaker might produce a nominative singular form **voisin* (on the model of the *pere* declension), *voisins* (on the model of the *murs* declension) or a form characterized by a different stem (***vois*?). This lack of informativeness in the case system constitutes, I suggest, a diachronic weakness which will be examined in further detail in chapter six.

Whilst in a syntagmatic approach instances of syncretism disrupt the canonical ideal of a direct form~function mapping, according to a WP perspective, syncretism can reduce uncertainty about the inflexion of previously unencountered nominals, sometimes providing minimally-entropic information. The distribution of sameness and difference in form in Romanian in particular is therefore very efficient - where a language-user knows two forms of a nominal (in addition to their functions), they can deduce with almost perfect certainty the rest of the paradigm. The distribution of fully-inflected wordforms in Romanian constitutes an instance of regularity which provides the language-user with predictive information about the form of a genitive-dative singular, nominative-accusative plural, or genitive-dative plural given knowledge of any form in the paradigm apart from the nominative-accusative singular. In old French and old Occitan, on the other hand, although the fully-inflected wordform is more informative than its sub-word parts and their distribution, the patterned distribution in

which these forms occur are not as unambiguous or informative as those of Romanian feminine nominals.

4.2.4 CLOSING REMARKS

The drawbacks of morpheme-based approaches have long been noted, particularly in reference to instances of fusional morphology. Where function and form are expected to be in a one-to-one mapping, a theory will often have to go to extraordinary lengths to manipulate forms found in a natural language to obey this isomorphic ideal. The data considered rarely fit without forcing the analyst to identify zero-morphs, empty morphs, and other 'solutions' which violate the spirit of a morphemic account. By their very nature, morpheme-based approaches cannot accommodate fusional languages such as Latin, Romanian, old French, and old Occitan without violating the basic notion of the morpheme as a 'unit'.

4.3.0 THE FINNIC DATA

In the following sections I will examine the evidence presented in chapter three and offer analyses of the Finnish, Livonian, and Veps inflexional case systems from constructivist and abstractivist approaches. Since it has previously been noted that morphemic models such as IA and IP are "optimized for the analysis of agglutinative patterns" in a language,[42] and since especially Finnish and Veps exhibit a great deal of agglutinative morphology, these morphemic accounts should in theory adequately accommodate the morphological structures found in these varieties. However, as will be demonstrated here, there are instances in which the form~function mapping is disturbed and which prove problematic to account for from a constructivist perspective. On consideration of these data, it will be concluded that WP is the model which can best accommodate the Finnic case systems, though it should be noted that fusional structure is less pervasive than in the Italic languages, as demonstrated by the more

[42] Blevins (2016:36).

frequent instances in which morphemic models can be used to describe the inflexional structures exhibited by particularly Finnish and Veps.

4.3.1 '(DE-)COMPOSITIONALITY' AND ATTENDANT ISSUES

In Finnish and in Veps in particular, cumulative expression of case and number is rare. Plural number in both these varieties, for example, is very consistently encoded by means of an easily identifiable and segmentable sub-word morph *-i-* which occurs between the lexical stem and the case suffix:

	Stem	**Number**	**Case**	**Result**
Inessive singular	*hebo-*	*-Ø-*	*-s*	*hebos*
Ablative singular	*hebo-*	*-Ø-*	*-lpäi*	*hebolpäi*
Translative singular	*hebo-*	*-Ø-*	*-ks*	*heboks*
Inessive plural	*hebo-*	*-i-*	*-š*	*heboiš*
Ablative plural	*hebo-*	*-i-*	*-lpäi*	*heboilpäi*
Translative singular	*hebo-*	*-i-*	*-kš*	*heboikš*

Table 4.16: Decomposing inflected forms of Veps {HEBO} 'horse'

In Veps, apart from the zero-marking of singular number, there are clear and easily-segmentable sub-word morphs which may be easily associated with particular functions. The plural marker is consistently *-i-* and the case values are expressed by the same suffixal shapes regardless of number (and, in fact, declension class).[43] These observations hold across the language.

However, there are a few instances, particularly in the grammatical case-forms of Finnish and Veps, where case and number values are cumulatively expressed. For example, in Veps, nominative plural functions are encoded by means of a form ending in *-d*; in Finnish by a form ending in *-t*. These 'nominative plural' suffixes are easily segmented away from the

[43] The palatalization of the sibilant in the inessive, elative, and translative forms following the high vowel *-i-* is regular throughout the written language.

fully-inflected nominal and it is uncontentious to describe them as inflexional material. However, there is no way to split the Veps form *-d* or the Finnish form *-t* into two morphs, one encoding 'plural number' and the other 'nominative case'. In order to preserve our morphemic analysis, we might posit a zero-morph encoding 'nominative', since this case value in the singular of these varieties is encoded by means of a lack of overt suffix, and conclude that *-d* is therefore a plural marker. However, throughout the rest of the plural in both Veps and Finnish, this function is encoded consistently by means of the morph *-i-*. It might be suggested instead that there are two plural allomorphs - *-d* and *-i-* in Veps and *-t* and *-i-* in Finnish - and that they are distributed according either to grammatical or phonological rules. For instance, the morphs *-t/-d* may be analysed as being distributed according to the grammatical environment [-oblique] and thus blocks the plural marker *-i-* which occurs elsewhere. This rule holds absolutely across the language, as the distribution of the allomorphs is the same for all nouns and adjectives.

In Livonian, the cumulative expression of case and (plural) number is much more common than in either Finnish or Veps, and thus there are many more instances in which the form~function mapping is disturbed:

	{KIK} 'rooster'	**{KE'Ž} 'hand'**	**{TU'Ļ} 'fire'**	**{KĪRAZ} 'axe'**
Nominative-genitive	*kikīd*	*kädūd*	*tu'ld*	*kirrõd*
Partitive	*kikīdi*	*ke'žži*	*tu'ļḍi*	*kirridi*
Dative	*kikīdõn*	*kädūdõn*	*tu'ldõn*	*kirrõdõn*
Comitative-instrumental	*kikīdõks*	*kädūdõks*	*tu'ldkõks*	*kirrõdõks*
Illative	*kikīž*	*kežīz*	*tuļīž*	*kirriž*
Inessive	*kikīs*	*kežīs*	*tu'ļši*	*kirris*
Elative	*kikīst*	*kežīst*	*tu'ļšti*	*kirrist*

Table 4.17: Cumulative morphs in Livonian plural forms

In the forms in table 4.17, it is extremely difficult to identify sub-word parts which can be consistently associated with particular feature values across all lexemes. It might be suggested that the high vowel *-ī-*, such as is found in the lexeme {KE'Ž}, for instance, is the plural marker, but it occurs only in the internal local case-forms; in all other forms (apart from the partitive plural) long *-ū-* is found instead, and the partitive plural is characterized by a distinct stem shape and a final *-(C)i*. Moreover, there is no sub-word exponent *-ī-* in the plural forms of {TU'Ļ} or {KĪRAZ}. Analysing *-ī-* as the plural morpheme therefore affords little generality across the language. We may instead recognize both *-ī-* and its short equivalent (found in certain forms in the plural paradigm of {KĪRAZ}, amongst others) as allomorphs of the plural morpheme. However, identifying grammatical or phonological environments which predict which of these forms is 'correct' is problematic, particularly when a larger dataset is taken into account.

Further issues arise on consideration of the grammatical case-forms (nominative/genitive, partitive, and dative) of all lexemes, the form *-d-* could be considered a plural allomorph alongside *-ī-/-i-*.[44] It could be argued that this allomorph is selected in the environment of 'grammatical' case values (as there are no phonological environments which can be analysed as licensing the selection of one allomorph over the other), as long as the translative-comitative is included with the grammatical case-forms (most traditional grammars list it as a 'semantic' case-form). The line between grammatical and semantic cases is notoriously difficult to draw (§1.2.2), and any arguments that are offered in order to make this grammatical environment predictive are likely to be arbitrary.

Finally, in the internal local case-forms of the lexemes in table 4.17, the high vowel which could be analysed as the plural morph usually occurs before the case ending (characterized by

[44] As long as the partitive plural form of {KE'Ž} '*ke'žži*' is analysed as *ke'ž+Di* where the dental *-D-* has assimilated with the preceding sound. We find similar assimilatory effects in the partitive plural form of {TU'Ļ}*: tu'ļḍi*).

a sibilant). However, in the plural inessive and elative of the lexeme {TU'Ļ}, this high vowel occurs *after* the case element. If we analyse *-ī-/-i-* as plural number markers, then other internal local forms in Livonian have the syntagmatic structure [STEM+no.+cx], the basic template which provides language-users with predictive information about the production of inflected forms in IA. However, in the forms *tu'ļši* and *tu'ļšti*, if *-i-* is analysed as the plural exponent, we have instead a structure [STEM+cx+no.]. There appears to be no phonological environment which selects this 'inverted' ordering of formatives, as can be demonstrated by comparing the inessive plural and elative plural forms of {TU'Ļ} with the same inflected forms of the lexeme {SU'Ž}. The latter nominal in the nominative singular is monosyllabic, characterized by a short medial *-u-*, the broken tone, and a final palatalized consonant and, in this way, exhibits very similar phonological shape to *tu'ļ*. However, {SU'Ž} exhibits inessive plural *sudīs*, and elative plural *sudīst*, following the structural template [STEM+no.+cx]. Why should two lexemes which exhibit such similar phonological structures exhibit distinct ordering of their morphemic constituents?

In order to solve such problems, diacritics may be introduced in order to provide information about the distinct sub-word structural patterns a nominal can follow. For instance: {TU'Ļ$_1$} in the lexicon belongs to the (arbitrarily named) 'first' class, which follow the structure [STEM+cx.+no], whilst {SU'Ž$_2$} belongs to the (arbitrarily named) 'second' class and follows the structure [STEM+no.cx]. Diacritic solutions, however, are often arbitrary, and, from an abstractivist perspective, 'flatten' the implicational structure exhibited by an inflexional system.[45]

In WP, such apparent alternations in the syntagmatic ordering of sub-word units are not problematic, since there is no requirement that sub-word elements, where they can be identified, must always be in the same order relative to one another in order. Instead, in

[45] Blevins (2006:544).

Livonian, more information about the structure of the inessive and elative case-forms in the plural is provided by other fully-inflected forms than by any sub-word morphs which may be identified or by 'schemata' for the syntagmatic distribution of these elements. In particular, the shape of either the inessive or elative in the plural usually provides minimally-entropic sufficient information to determine the structure of the remaining internal local case-form.[46] For instance, if the inessive singular is characterized by a form *-sõ*, the elative singular will end in *-stõ*; if the elative singular ends in *-õst*, the inessive singular will end in *-õs*; similarly, if we find a plural form *-si* in the inessive we can expect *-sti* in the plural and vice versa; inessive plural *-is* and *-īs* imply elative plural *-ist* and *-īst* respectively. Thus, the inessive singular *tu'ļši* implies elative singular *tu'ļšti*, whilst inessive plural *sudīs* implies elative plural *sudīst*. Splitting these forms reduces the amount of information provided by the fully-inflected shapes and the abstract relations that exist between other forms in the paradigm.

4.3.2 ALLOMORPHY

Phonological variation of suffixal case markers across number and/or declension is relatively rare in the inflexional case systems of the Finnic languages. However, at least one of the grammatical case-forms in each of the varieties discussed in chapter three exhibits non-phonologically-conditioned formal variation across declension; most grammatical case-forms exhibit non-phonologically-conditioned formal variation across number. Livonian stands out as a Finnic variety which exhibits distinct exponents for almost all productive morphological case-forms across many declensions and most numbers.

One striking exception to the above generalizations is that of the morphological variation exhibited by the illative case suffix in Veps. As stated in chapter three, there are nine desinences associated with the illative: *-he*, *-ha*, *-ho*, *-hu*, *-hä*, *-hö*, *-hy*, *-ze*, and *-že*. From an

[46] Notably, the illative plural case form does not always follow the same sub-word structure as the other inessive and elative, as seen in the paradigm for {TU'ļ} in table 4.17.

IA perspective, it might be argued that the lexical stem of a nominal would be stored, as would the illative shape *-hV*, the final vowel of which, when combined with the nominal stem, takes on the quality of that which precedes it; *-ze* always follows a stem in *-hV-* (e.g., {ROBEH} 'birchbark', common stem *robehe-*, illative singular *robehe-ze*) whilst *-že* follows a stem in *-hi-* (e.g., {VERAS} 'stranger', plural stem *verhi-*, illative plural *verhi-že*). A phonologically-conditional approach therefore seems, at first blush, to accommodate this extensive allomorphy.

However, many nominals exhibit stems in the illative singular form which are characterized by a final consonant, to which the illative singular form *-hV* 'attaches'. It is not clear on consideration of these forms how to determine the nature of the final vowel of the illative desinence with reference only to the phonological environment:

		-he		***-ha***	
Stem ending in:	**(polysyllable) + vowel**	*kodi-he*	(in)to home	*kala-ha*	into the fish
	Consonant	*näp-he*	into the pinch	*jodog-ha*	into the bucket
	Consonant cluster	*ong-he*	into the fishing rod	*adr-ha*	into the plough
	(C)V	*te-he*	into the road	*ma-ha*	into the country/ on(to) the ground

Table 4.18: Illative singular forms in written Veps

How do speakers determine that, for a nominal such as {ADR}, the form *-ha* should follow the consonant cluster, but that *-he* should follow the consonant cluster found in the stem shape *ong-*? A wider range of examples demonstrates that no phonological conditioning environments can be found to determine the shape of the following illative singular exponent. For instance, both *onghe* and *jodogha* exhibit an *-o-* in the syllable preceding the illative singular suffix, and yet the quality of the final vowel of the suffix differs. The only difference is that the form *-he* which, in the first instance, is preceded by a consonant cluster *-ng* but that

-ha follows *-g*. However, there is an illative singular form *bongha* '(in)to the maelstrom', where the form *-ha* is preceded by the consonant cluster *-ng-* as well as by a vowel *-o-*.

In order to correctly form the illative singular of certain forms in Veps, reference to the paradigm is required. An account which recognizes the central importance of the paradigm is therefore the only model which can accommodate the formal variation of the Veps illative singular exponents. For instance, it is not clear how the illative singular forms in table 4.19 are meant to be derived by knowing only the relevant stem shape (here identical to the nominative singular form) of the nominal:

	{TAT} 'father'	**{KAIK} 'all'**	**{LIND} 'bird'**	**{MAID} 'milk'**
Nominative	*tat*	*kaik*	*lind*	*maid*
Illative	*tatha*	*kaikhe*	*lindhu*	*maidho*

Table 4.19: Nominative and illative singular forms of selected Veps nominals

Although the form *tatha* might be considered transparent (where it may be suggested that *-ha* is selected because the vowel in the stem is also *-a*), it is not clear why *-hu* should be selected for the nominal *lind* or *-ho* for *maid*. However, consideration of other inflected forms in the paradigm shed light on the formation of the illative singular:

	***tat* 'father'**	***kaik* 'all'**	***lind* 'bird'**	***maid* 'milk'**
Nominative	*tat*	*kaik*	*lind*	*maid*
Genitive	*tatan*	*kaiken*	*lindun*	*maidon*
Partitive	*tatad*	*kaiked*	*lindud*	*maidod*
Illative	*tatha*	*kaikhe*	*lindhu*	*maidho*

Table 4.20: Partial paradigms of selected Veps nominals

It is clear, when considering other inflected forms of nominals, that the vowel quality of the illative singular desinence *-hV* is determined by the nature of the vowel which occurs between the nominal stem and the inflexional exponent elsewhere in the singular paradigm. For instance, in the genitive singular form *lindun*, *-u-* intervenes between the stem and the genitive singular exponent *-n*. This stem vowel is not present in the illative singular stem

form, but is nevertheless the vowel that is used in the illative singular desinence *-hV*. Knowledge of other inflected forms of a lexeme is therefore essential in determining the shape of the illative singular, particularly for nominals with a nominative singular ending in a consonant.[47] Similar observations hold for the terminative, since this case form has its origins in a postposition which occurred with a form inflected in the illative. Since this 'illative' element is what introduces formal variation in the terminative case-form across number and declension, the explanation given above for the illative singular also accommodates the distinctions exhibited by the distinct terminative exponents. This may also lend paradigmatic support to the extensive inflexional variation exhibited by the terminative suffixal forms.

In Livonian, all productive inflexional case values apart the nominative singular and genitive singular have at least two distinct exponents. Some of these, such as the dative singular, may be described as being selected according to the phonological environment in which they occur: *-n* follows a stem ending in a vowel, *-õn* occurs with a stem ending in a consonant. However, other forms are not so easily to account for from a morpheme-based account.[48]

Despite frequently-opacified form~function mapping, there are often syntagmatic patterns which may be recognized and exploited by speakers in the production of novel inflected forms in Livonian. An abstractivist account can capture sub-word 'regularities' without requiring that forms be decomposed and analysed into meaningful building blocks. Moreover, it is possible to offer observations that hold across declensions, even where nominals exhibit distinct inflexional behaviour, as the patterns WP recognizes as potentially informative are not restricted to the paradigm. Some of the most wide-spread generalizations are as follows:

[47] The formation of the illative is not actually as simple as presented here. Ainsworth (2019a) discusses the formation of the Veps illative from both constructivist and abstractivist accounts.
[48] Discussed in reference to the inessive and elative forms in §4.3.1.

1. the dative singular form is always characterized by a final *-(õ)n*;
2. the nominative plural and genitive plural functions are always encoded by the same form;
3. the nominative/genitive plural form is always characterized by a final *-D*: *-t* follows a sibilant and *-d* occurs elsewhere;
4. the dative plural form is always characterized by a final form *-õn* which occurs with a stem the same shape as the nominative/genitive plural form;
5. the translative-comitative form always ends in *-(õ)ks* or *-kõks* (the distribution of these forms is morphologically determined);
6. the partitive plural form always ends in *-(D)i*;
7. in the plural, the sibilants found in the internal local case-forms are palatalized, except where another palatalized sibilant precedes the inflexional ending;
8. the nominative singular and genitive singular forms of a lexeme are almost invariably characterized by the same number of syllables;[49]
9. where a partitive singular form ends in *-õ* (not *-Dõ*), the illative singular form will be identical;
10. where the genitive singular form ends in a final *-õ* and the illative singular ends in *-õ*, the illative singular form can take an optional final *-z*;
11. the elative form can always be deduced from the inessive form by the addition of a *-t-* immediately following the sibilant of the inessive form (in both singular and plural);
12. similarly, the inessive form can always be deduced from the elative form by the omission of the *-t-* immediately following the *-s/-š* of the elative form (in both singular and plural).

[49] Ainsworth (2019b:16) finds only two inflexion classes which violate this generalization: nominative singular ~ genitive singular *neitst ~ neists/neitsõ* 'virgin, maid' and *sīend ~ sīend/sīenõ* 'full, having eaten'. These are populated by only one or two members according to the Livonian dictionary online.

These implicational relations are not recognized in morphemic models of grammatical description, in which it is the relations below the level of the word which are considered informative in the production of novel inflected forms. However, it will be noted that most of the generalizations 1-12 are offered about sub-word parts - that is, stem and suffixal shapes and their distribution within the paradigm and across the language. WP, of course, allows for the recognition of sub-word recurrent partials as units of analysis, particularly where they provide a language-user with information about the production of novel forms.

Allomorphy in the lexical stem is more pervasive than phonological variation in case suffix in the Finnic languages. As has already been discussed, certain instances of stem allomorphy can be problematic for morphemic models. In Finnish, there are certain instances where the strict concatenation of lexical stem and inflexional suffix (or suffixes) produces the correct surface forms:

Stem	**Number**	**Case**	**Concatenation**	**Output (written)**
{KUIVA}	{PLURAL}	{ABLATIVE}	/kwivʌ-i-ltʌ/	**kuivailta*
{KUIVA}	{PLURAL}	{ESSIVE}	/kwivʌ-i-nʌ/	**kuivaina*
{KUIVA}	{PLURAL}	{INESSIVE}	/kwivʌ-i-sːʌ/	**kuivaissa*

Table 4.21: Incorrect plural forms of Finnish {KUIVA} ‘dry’[50]

In the modern language, there is no restriction on the combination of the vowels /ʌ/ and /i/ (the diphthong /ʌi/ is relatively common in lexical and inflected forms, e.g., *n<u>a</u>i<u>n</u>en* ‘woman’ *vanh<u>a</u>-i<u>n</u>*, a genitive plural form of {VANHA} ‘old’). For this reason, we cannot suggest that this diphthong does not occur in surface forms because of the restriction of the co-occurrence of these phonemes. Nevertheless, in this lexeme type (and, in fact, in many nominal types across the language), we find a different stem shape in all semantic forms in the plural when compared to the semantic forms in the singular.

[50] The correct output forms would be *kuivilta*, *kuivina*, and *kuivissa* respectively.

If all plural forms were characterized by the same stem, we could argue that a certain stem shape is 'selected' either by the grammatical environment [+plural] or by the phonological environment 'before *-i*'. However, the nominative plural in Finnish (and, in fact, all Finnic varieties) is almost always characterized by a stem shape distinct from the other plural case-forms.[51] Here, therefore, a grammatical analysis falls down. A phonological approach may be suggested instead, where distinct stem allomorphs are analysed as being 'selected' when followed by the plural form *-i-*. This accounts for the distinct stem forms in the nominative plural compared to the rest of the plural forms in a paradigm, as the nominative plural form is not characterized by a plural marker *-i-*. However, on comparison of lexemes with a common stem in *-A* (e.g., {SUKKA} 'sock', {OMENA} 'apple'), the phonological analysis becomes unsatisfactory, as before *-i-* certain forms lose the final *-A* of the common stem, and in others, the *-A* becomes *-O* (e.g., *suka-ssa* inessive singular, *suk-i-ssa* inessive plural; *omena-ssa* inessive singular, *omeno-i-ssa* inessive plural). Without resorting to diacritics (the problems associated with which have been briefly discussed in §4.3.1 in reference to Livonian), there is no way to determine whether an *A*-stem nominal will follow the pattern of *sukka-* or *omena-*type nominals.

Stem allomorphy is even more pervasive in Livonian than in Finnish. For the most part, stem alternations are the result of regular phonological developments, but synchronically it is difficult to account for the distinct shapes of nominal stems and their distribution. For instance, where the genitive singular form is formally distinct from the nominative singular, it is generally differentiated by means of morphophonological differences in the nominal stem (rather than by overt suffixation, as in all other forms in the Livonian inflexional paradigm):

[51] 'Almost always' because there are a few stems which are invariant within the inflexional paradigm, e.g. {NALLE} 'teddy', {TALO} 'house'.

Nominative singular	Genitive singular	Gloss
me'r	*mie'r*	sea
kiļg	*kilg*	side, aspect
rikāz	*rikkõ*	rich
pǭļaz	*paļļõ*	naked, bare
pu'nni	*punīz*	red
läpš	*laps*	child

Table 4.22: Morphophonological alternations in Livonian nominal stems

The differences given here (which are certainly not exhaustive of the alternations found across the language but which are representative of the types of distinctions which occur) include: alternations in tone (*pu'nni ~ punīz*), in consonant length (*rikāz ~ rikkõ*), in consonant quality (*kiļg ~ kilg*), the length and/or quality of the 'stem' vowel (*pǭļaz ~ paļļõ*; *läpš ~ laps*), or in a mixture of these. In none of these examples is the genitive singular form distinguished from the nominative singular form by means of phonologically invariant affixal material. How do we account for the many instances, and different types, of formal variation found in the lexemes in table 4.22?

In the lexemes {KIĻG} and {LÄPŠ}, the palatalization of the *-l-* and *-s* in these forms might be analysed as being introduced by a final high vowel *-i* which is underlyingly present but which is not pronounced.[52] The alternation of the medial vowel *ä ~ a* in *läpš ~ laps* may be analysed similarly as occurring where the stem is followed by *-i* which 'fronts' the *-a-* found in most other forms, which also accounts for the partitive and internal local case-forms in the plural. Whether or not it reflects the ways in which language-users actually deduce these forms is, of course, a matter for debate, particularly because the reality of this hypothesis can no longer be tested on native speakers.

Not all the alternations found in nominative singular and genitive singular forms may be easily explained by suggesting that an underlying final high vowel caused instances of

[52] Historically, these nominative singular forms would have ended in *-i*. The suggestion put forward here therefore reflects the etymology of these forms.

consonant palatalization or vowel fronting, however. For instance, the forms *pǭļaz ~ paļļõ* and *rikāz ~ rikkõ* show similarities in structure, where the nominative singular forms are both characterized by a final *-z* and a bisyllabic stem, and the (bisyllabic) genitive singular stem ends in a final *-õ* with a lengthened medial stem consonant. But it is not clear how an underlying final *-i* should account for these alternations. Instead, we might suggest that there is another nominative singular morph *-z*,[53] which causes consonant gradation (since it closes an open syllable which is found in the genitive singular forms), and which therefore accounts at least for the consonant alternations found in these forms.

However, there are also alternations in the length and sometimes also quality of the vowels found in the stem of these lexemes. For instance, the vowel in the first syllable of *pǭļaz* is *-ǭ-* but changes to *-a-* in the same position in the genitive singular form; in *rikāz*, it is the vowel in the second syllable which alternates, shortening and becoming *-õ*. If we suggest that the genitive singular form may be derived by removing the final *-z* from the nominative singular form (with subsequent lengthening of stem-medial consonants), then it is not clear why its loss leads to the shortening of the vowel in the first syllable in one lexeme, but the shortening of the vowel in the second syllable in another.

Even though many of the stem alternations found in the Finnic languages cannot be derived via synchronically-regular phonological processes, there is evidence to suggest that the alternations exhibited by different stem shapes within the inflexional paradigm of a lexeme are nevertheless recognized and exploited in language-use. For instance, there is some fluctuation in the morphological behaviour of stems of *ajatus*-type nominals and *taivas*-type nominals in Finnish:

[53] Setting aside for the moment problems associated with allomorphy.

{AJATUS} 'thought' (*-Vs ~ -Vks-*)		{TAIVAS} 'heaven' (*-Vs ~ -VV-*)	
Basic form	**Inflexional form**	**Basic form**	**Inflexional form**
ajatus	*ajatukse-*	*taivas*	*taivaa-*
kysymys	*kysymykse-*	*valmis*	*valmii-*
vastaus	*vastaukse-*	*rikas*	*rikkaa-*
teos	*teokse-*	*oppilas*	*oppilaa-*
rakennus	*rakennukse-*	*tehdas*	*tehtaa-*
hallitus	*hallitukse-*	*kirves*	*kirvee-*
päätös	*päätökse-*	*kirkas*	*kirkkaa-*

Table 4.23: Stem differences in *ajatus*- and *taivas*-type nominals

Judging from these examples, it can be argued that nominals with a basic form ending in *-Us*, or *-Os* belong to *ajatus*-type nominals, and those ending in *-As, -is* and *-es* belong to *taivas*-nominals. But there is synchronic evidence to suggest that speakers no longer associate these nominal types as being determined by the shape of the nominative singular in all lexemes. Many lexemes which are etymologically characterized by the alternation *-Vs ~ -VV-* now follow the inflexional pattern *-Vs ~ Vks-*. In fact, at least two lexemes {UROS} 'man; male' and {KOIRAS} 'male' can decline like either {AJATUS} or {TAIVAS}:

ajatus-type	**Singular**	**Plural**
Nominative	*uros*	*urokset*
Genitive	*uroksen*	*urosten, uroksien*
Partitive	*urosta*	*uroksia*
Illative	*urokseen*	*uroksiin*
taivas-type	**Singular**	**Plural**
Nominative	*uros*	*uroot*
Genitive	*uroon*	*uroiden, uroitten*
Partitive	*urosta*	*uroita*
Illative	*urooseen*	*uroisiin*

Table 4.24: {UROS} 'man' partially declined

In the modern language, it appears as though nominals originally belonging to the same declension as {TAIVAS} can be reinterpreted as belonging to the same class as {AJATUS}. This highlights the occasional weaknesses contingent on a pedagogical abstractivist approach which aims to identify declension classes which exhaust all or the majority of the inflexional

variation found in the language. By taking a broader, descriptive (as opposed to prescriptive) view, as is possible in WP, we can account for the formal variation found in the language which is not necessarily captured in didactic grammars.

When we are not concerned with the derivation of stem allomorphs from an underlying form, as in IA or IP, and instead take a broader view of the data, it becomes clear that, although the shape of stems is not always predictable, the distribution of different forms within the paradigm often is. In Veps, for instance, the nominative singular is frequently characterized by a slightly different stem shape when compared to the majority of the rest of the paradigm; the partitive singular, too, sometimes has a distinct phonological shape; finally, there is occasionally a slightly different stem shape for the illative. The rest of the singular forms, however, as well as the nominative plural, are characterized by the same common stem. The plural forms (apart from the nominative), in all nominal types but *i*-stems, are marked by a different stem form:

	Singular	**Plural**
Nominative	NOMINATIVE STEM	COMMON STEM
Genitive	COMMON STEM	PLURAL STEM
Partitive	PARTITIVE STEM	PLURAL STEM
Ablative	COMMON STEM	PLURAL STEM
Adessive	COMMON STEM	PLURAL STEM
Allative	COMMON STEM	PLURAL STEM
Elative	COMMON STEM	PLURAL STEM
Inessive	COMMON STEM	PLURAL STEM
Illative	ILLATIVE STEM	PLURAL STEM
Egressive	COMMON STEM	PLURAL STEM
Approximative	COMMON STEM	PLURAL STEM
Propinquative	COMMON STEM	PLURAL STEM
Comitative	COMMON STEM	PLURAL STEM
Prolative	COMMON STEM	PLURAL STEM
Terminative	COMMON STEM	PLURAL STEM

Table 4.25: Stem allomorphy distribution in Veps nominals

This distribution may be described as morphomic, as there are no semantic, syntactic, or phonological considerations which can accord for the common stem occurring in the nominative plural cell, for instance. However, despite the 'unprincipled' distribution of these stem shapes, since the pattern is complete throughout the nominal system, it may be analysed as one which can be recognized and utilized by language-users in the production of novel inflected forms. Certainly, the occurrence of the common stem in the nominative plural form of a lexeme is consistent across the Finnic languages, which may provide diachronic evidence which supports the hypothesis that the patterned distribution is one to which language-users are sensitive. Such patterns are best captured by WP in which the paradigm as a whole and the relations between inflected forms are highly informative.

4.3.3 SYNCRETISM

Instances of syncretism are less frequent in the Finnic languages than in the Italic varieties, though they do occur. One example is the comitative in Finnish. This form is easily segmentable into a stem, a number morph, and a case morph:

Comitative plural: *kuiv-i-ne*

Stem	number	case
{KUIVA}	+ {i}	+ {ne}
/kwiva/	+ /i/	+ /ne/
[kuiv]	+ [i]	+ [ne]

Output form: *kuivine* (comitative plural)

Figure 4.6: Decomposing Finnish comitative *kuivine* 'dry'

But there is no way of correctly generating the correct comitative singular form *kuivine* by concatenating the singular stem, number marker -*Ø* and case morph -*ne* (which would give us **kuivane*). Instead, we might derive the comitative singular form by adding a process morph which applies an operation which 'maps' the comitative singular functions onto a form which happens to be identical to that of the comitative plural. Since this syncretism is absolute throughout the language, this process holds across all nominals. Similar observations hold for the instrumental and prolative forms.[54] Once again, however, this suggestion does not 'solve' the problem, for a morphemic account, that a single form encodes multiple different sets of functions and, where instances of syncretism are not absolute, such a solution provides little generality across a language.

According to a WP analysis, the systematic pattern of syncretism provides a language-user with minimally-entropic information about the forms of the comitative, instrumental, and prolative in Finnish. Thus, where a language-user knows the comitative plural form, they also know the form of the comitative singular (and vice versa); knowing the prolative plural form provides minimally-entropic information about the prolative singular, etc.. That is, rather than constituting an instance of 'irregularity' that requires special and sometimes *ad hoc* explanations for its occurrence in language, syncretism can actually help to reduce the entropy of an unknown form. It is also notable that this pattern occurs in three of the least frequent case-forms found in Finnish. Thus, rather than having to deduce six distinct forms for values which are rarely used, language-users may make use of half that number. From this perspective, syncretism in Finnish may be considered a 'facilitating' aspect of the case system, (rather than an opacification in the sub-word form~function mapping, as in a morphemic account) which aids in the production of novel forms.

[54] The syncretism of the instructive is similarly pervasive, although there is a single instance of which I am aware where the systematic syncretism of the instructive across number is interrupted: the noun {JALKA} 'foot' has a singular form *jalan* and a plural form *jaloin*.

Syncretism is more pervasive in Livonian than in Finnish or Veps. The nominative plural and genitive plural functions are always encoded by a single form, as are the translative and comitative functions in the singular and (by a different form) in the plural.[55] The syncretism of the nominative and genitive functions in the singular, on the other hand, is not absolute and may therefore be described as 'accidental'. Where a lexeme exhibits this pattern of syncretism, an IP account may apply a process of no operation to the stem which means that it encodes 'genitive' as opposed to 'nominative' functions. Knowing which lexemes to which to apply this operation, however, is tricky, since there are no generalizations that can be made about the semantic content of these lexemes or their morphological structure, for instance; the occurrence of syncretism is something that must be learned on a lexeme-by-lexeme basis. No generality is offered by an IP account of this pattern of syncretism in Livonian.

Once again, however, it is noted that WP does not necessitate the identification of maximally-predictive phonological, grammatically, or semantic environments in which patterns of syncretism (such as that of the nominative and genitive in the singular) do or do not occur. Language-users may certainly make use of patterns which reduce the uncertainty about the shape of a novel form, and, since an abstractivist account does not prioritize only paradigmatic patterns in the production of novel forms, other clues may be identified that help to reduce the uncertainty associated with producing a particular novel inflected form.

Gloss:	**rooster**	**fish**	**top**	**hand**	**fire**	**hot**	**woman**	**axe**
Nom.sing.	*kik*	*kalā*	*ladā*	*ke'ž*	*tu'l*	*tu'ļļi*	*nai*	*kīraz*
Gen.sing.	*kik*	*kalā*	*ladā*	*kä'd*	*tu'l*	*tuļīz*	*naiz*	*kirrõ*

Table 4.26: Nominative singular and genitive singular forms of Livonian nominals[56]

[55] In Courland Livonian that there are a few nominals which do differentiate the comitative and translative morphologically. This complicates the picture, as IP can more easily account for systematic syncretism (as in the Finnish comitative) than for apparently sporadic exceptions.
[56] Adapted from Viitso (2007:52).

Four of the eight lexemes given in table 4.26 do not distinguish ‘nominative singular’ from ‘genitive singular’ functions morphologically. That is, in 50% of instances, knowledge of the nominative singular form predicts absolutely the shape of the genitive singular form. In the other lexemes, this is not possible. As a result, it is equally likely that: a) the genitive singular form will be different from the nominative singular, and that b) the two functions will be encoded by the same form. Thus, if a speaker wants to produce a genitive singular the same shape as the nominative singular, they are likely to guess the correct form only half of the time.[57]

There are, however, other patterns which may be utilized in reducing the uncertainty associated with determining whether ‘nominative’ and ‘genitive’ functions in the singular are encoded by the same form. For instance, the phonological structure of certain forms may provide information about the declension class to which a nominal belongs, and therefore about the nominative singular and genitive singular forms. For instance, on consideration of nominals in the Livonian dictionary, lexemes with a (bisyllabic) nominative singular form ending in *-a/-ā* all have an identical genitive singular form. In this way, although there is, according to the assessment given above, a 50% chance of correctly guessing whether the nominative singular and genitive singular forms will be distinct in the singular of the lexemes in table 4.26, this uncertainty is eliminated for *-a/-ā* nominals by reference to other lexical items in the language with a similar phonological structure. This is not to say, of course, that all nominals with a nominative singular form ending in *-a/-ā* follow exactly the same pattern (§3.2.1). But knowledge of the nominative singular form, as well as reference to the inflexional paradigms of other forms with similar phonological structure, can provide information about the templatic distribution of inflected forms.

[57] Although, of course, this calculation and the discussion which follows does not take into account the fact that these declensions: a) do not include all the inflexional variation found in all the declensions identified in the Livonian dictionary; b) are not populated by the same number of lexemes, and c) certain inflected forms are likely to occur more frequently than others.

4.3.4 CLOSING REMARKS

Particularly in the Finnish and Veps case systems, there are many invariant sub-word elements which may be associated with particular functions in a direct and transparent way. The plural is consistently marked in Finnish and Veps in all case-forms but the nominative, and particularly the semantic case markers are generally invariant across both number and declension class; instances of allomorphy in the stem tend to be synchronically-transparent, as do phonological distinctions between case exponents across number and/or declension. The syntagmatic structure of inflected nominals in these varieties is a lot less opaque than that found in Latin, old Gallo-Romance, and Romanian.

Despite the frequent sub-word clarity of these case systems, however, morphemic accounts frequently cannot accommodate certain morphological structures found in each of the varieties considered here. Abstractivist accounts, on the other hand, can recognize sub-word parts which may be abstracted away from fully-inflected forms and associated with particular functions, but do not necessitate that form and function be in an isomorphic mapping below the level of the word. In this way, WP can accommodate the more fusional structure of particularly Livonian as well as the more agglutinative structure of Finnish and Veps.

4.4 CONSTRUCTIVIST AND ABSTRACTIVIST APPROACHES - CLOSING REMARKS

Morphemic models cannot easily describe some of the instances of fusional structure found in the morphological case system of both the Italic and Finnic languages, and even in languages with more agglutinative morphological structure it is rare to find a case system where all fully-inflected wordforms can be decomposed into sub-word units which can be consistently associated with particular functions. An abstractivist approach certainly best accommodates all the languages discussed in this study.

However, unlike a constructivist account, an abstractivist account does not afford importance to sub-word recurrent partials. By considering the Italic and Finnic data from morphemic approaches, it has become clear that there is a striking difference in the morphological structures which characterize these languages. That is: Finnish and Veps in particular exhibit an abundance of agglutinative structure; Latin, old Gallo-Romance, Romanian, and Livonian exhibit more fusional structure. Whilst WP can accommodate these different morphological structures, it does not recognize a qualitative difference between instances in which form and function are in a biunique relationship and those in which this direct mapping is interrupted in some way. As will be explored in more detail in the next chapter, there is comparative historical evidence to suggest that these two types of morphological structures behave different in synchrony and diachrony. In chapter six, it will be argued that these different morphological structures shed light on the developments of the Italic and Finnic case systems.

5.

‘SUBSTANCE ENTITIES’ AND ‘SCHEMATIC PATTERNS’

5.0 INTRODUCTION

In the following sections, I will present evidence for the claim made at the end of the previous chapter, namely: that a qualitative difference between certain types of sub-word inflexional variation must be recognized in morphological theory. This can be seen most clearly a) synchronically, in the ways morphological alternations are recognized and utilized by language-users in the production of previously unencountered inflected forms, and b) diachronically, in the ways that they are likely to develop through time. I suggest that there is an important distinction to be made between morphological structure which is in a direct form~function mapping and that in which the relationship between form and function is disturbed in some way. Evidence for such a claim will be outlined in depth below, following the explanation of the differences between what I call ‘substance entities’ and ‘schematic patterns’.

The claim made here may seem at first blush to accord with that of Lieber (1980), Noyer (1992), Halle and Marantz (1993), and Carstairs-McCarthy (1994), who argue that there is “an important difference between affixal and nonaffixal (or between concatenative and nonconcatenative) morphology in respect of what they do or how they are constrained or both”.[1] The opposing view is that “[t]here is no theoretically significant difference between concatenative and nonconcatenative inflection”,[2] and is supported by Lieber (1988, 1992), Anderson (1992), Stump (2001), and Blevins (2004).

The second view is adopted here, as the arguments in its favour are more persuasive than those offered against it. For instance, as Blevins discusses in reference to Carstairs-McCarthy’s notions of ‘Paradigm Economy’ and the ‘No-Blur Principle’, a comparison of a

[1] Carstairs-McCarthy (1994:739).
[2] Stump (2001:9).

Word and Paradigm (WP) approach and one in which forms are necessarily split into 'stem' and 'affix' "suggests that generalizations over affixal exponents are derivative of patterns of interdependence involving whole words, and, hence, that there is ultimately no need for dedicated stem- or affix-based economy principles."[3] Furthermore, the reasoning for a distinction between affixal and non-affixal morphology given in Carstairs-McCarthy's (1994) paper feels rather circular: he at first assumes to distinguish suffixes from stems in order to put forward his 'No-Blur' principle, and later uses the 'evidence' discussed in reference to proving that such a restriction exists to validate his decision to ignore non-affixal morphology.[4] Stump provides a number of counter-examples to the No-Blur Principle from Vedic Sanskrit and concludes that the No-Blur Principle cannot disconfirm the assumption that there is no theoretical difference between concatenative and nonconcatenative morphology because the Principle itself is not valid.[5]

Rather than suggesting that concatenative and nonconcatenative morphology should be treated as somehow 'theoretically distinct', the view espoused here is that there is a difference between morphological elements which stand in a one-to-one form~function mapping and instances in which sub-word morphological patterns cannot be easily extricated from the speech stream and associated with particular functions, and the ways they are exploited in language use. This chapter will focus on elucidating and providing acquisition and comparative historical data to lend support to this hypothesis.

---

[3] 2004:41.

[4] "[I]n short, restricting attention to affixes is useful, because affixal exponents can - at least usually - be isolated, and convenient, because it provides a basis for ignoring patterns that would otherwise violate economy principles." Blevins (2004:54).

[5] 2005:292. There does appear to be, however, a potential difference between the way concatenantive and nonconcatenative morphology are treated in very early on in language acquisition, where "[c]ase-marking in agglutinating languages occurs very early: the first case oppositions in the speech of a Turkish child were observed at 1;3, and two Finnish children were found to utter their first case-forms in appropriate contexts even earlier (at 1;0-1;3) [...]. Children speaking Slavonic and Baltic languages start with case oppositions at a relative later age: the mean age of the first occurrences is about 1;9 [...]. In languages with periphrastic case-marking on nouns and articles, the first contrastive forms emerge at the latest, between 2;2 and 2;3". Eisenbeiss et al. (2011:381f.). This sort of evidence suggests that language-users recognize and utilize sub-word clarity in the deduction of a novel form very early on. However, as the child grows, it appears as though concatenative and non-concatenative morphological structure are used and understood with equal ease.

5.1.0 ‘SUBSTANCE ENTITIES’ AND ‘SCHEMATIC PATTERNS’

This section will introduce and explain what is meant by ‘substance entities’ and ‘schematic patterns’. These terms are employed here in reference to two different types of morphological variation: that which has to do with forms which can be consistently associated with a particular function or set of functions, and that which has to do with inflexional variation which cannot easily be extricated away from the speech stream and associated with particular functions, but which may constitute informative patterns which aid in the deduction of novel inflected forms.

The terms ‘substance’ and ‘schematic’ are adapted from discussion on contact-induced language change in Croft’s (2000:201-205). In his discussion on this topic, ‘substance’ interference refers to the borrowing of chunks of phonological material which can be associated with particular meanings and which are transplanted wholesale from one language to another, e.g., the loaning of lexical items. ‘Schematic’ borrowings, on the other hand, have to do with the interference of abstract patterns, such as the influence of word order on a contact language (which might help induce or speed along a change from SOV to SVO, for instance). Drawing from this discussion, I employ the terms ‘substance entity’ and ‘schematic pattern’ here in the following ways:

- ‘Substance entity’ - a continuous segment of phonological material that can be extricated from the speech-stream and associated with a particular function or set of functions. Here, I will be concerned with two different types of substance entity: ‘subword’ and ‘word-level’. Subword substance entities can consist of (largely) invariant case and number exponents; the term ‘word-level substance entity’ refers to a fully-inflected wordform. Where a sub-word form exhibits transparent formal variation (for instance, the Finnish partitive, the forms of which are phonologically

very similar and which are all easily associated with the same functions), it will be considered a substance entity. Substance entities above the level of the word - such as periphrastic forms - will not be discussed.

- 'Schematic pattern' - this term refers to patterns of formal variation that exist between inflected forms of a lexeme which are *not* in an isomorphic form~function mapping.[6] Unlike substance entities, schematic patterns cannot be extricated from the speech stream whole and associated with particular (sets of) functions. By their very nature, schematic patterns and the functional distinctions they encode can only be described by directly comparing two or more forms. Common patterns of alternation cross-linguistically include 'ablaut', 'umlaut', and 'consonant gradation'. Instances in which affixal morphology where function and form are not in a one-to-one mapping, such as the case/number suffixes in Latin, are also considered to be examples of schematic patterns since, although these sub-word forms can often be segmented away from fully-inflected wordforms, they cannot be consistently associated with particular functions.

The following discussion will focus on presenting examples and further explaining these two types of morphological variation, which will be illustrated using the following example:

The cats and the geese eat some food

Here, the forms 'the', 'cat', '-s', 'cats', 'and', 'the', 'geese', 'eat', 'some', and 'food' can all be described as substance entities. (Note that the constituents 'the cat' and 'the geese' are not considered substance entities, since the definite article and noun can be 'interrupted' by, for instance, an adjective 'cute', 'angry' etc..) These are forms which can occur in other

[6] It should be stressed here that the term 'isomorphic' is used throughout this and the following chapters not to refer necessarily to a 'one-to-one' mapping of form and function, but, frequently, to a direct mapping of form and a set of functions. That is, in Finnish a final *-t* can be associated with the feature value set {CASE:NOMINATIVE; NUMBER:PLURAL}.

contexts, in combination with other lexemes, but which can be consistently be associated with specific functions:

- ‘cat’ - feline;
- ‘-s’ - in combination with a noun, often and reliably associated with the function ‘plural’;[7]
- ‘cats’ - multiple felines;
- ‘food’ - sustenance consumed by people or animals for life and growth; etc.

These forms can only be identified as substance entities by reference to and knowledge of other forms in the language. Speakers will encounter the forms ‘the’, ‘cat’, ‘eat’, ‘some’, and ‘food’ in other contexts, which will enable them to recognize common functions for which these signs stand. From the many instances in which a language-user meets the substance entity ‘cat’, they can build up a sense of what ‘meaning’ this word has:[8]

the cat is purring

she is stroking a cat

*‘Cat’: “a well-known carnivorous quadruped (*Felis domesticus*) which has long been domesticated, being kept to destroy mice, and as a house pet”*[9]

the dog chases the ginger cat

etc.

[7] Although a genitive clitic /s/ can occur with nouns, in this syntactic context the final *-s* must be analysed as the plural suffix. Recent psychoacoustic studies have indicated that speakers are actually sensitive to sub-phonemic contrasts between seemingly homophonous forms, including plural *-s* and genitive clitic *-s*. Gahl (2008), Drager (2011), Plag et al. (2017), Baayen et al. (2003), and Kemps et al. (2005).

[8] This reflects the ‘encyclopaedic approach’ to linguistic meaning as already discussed in §1.1.2.

[9] http://www.oed.com/view/Entry/28649?rskey=6BTvdg&result=2&isAdvanced=false#eid. Accessed: 14/04/2019, 10:10. It is unlikely that native speakers have this definition in mind for the lexeme {CAT}. This entry is instead offered here to represent the fact that L2 learners in particular may sometimes consult a dictionary rather than deduce the function(s) of a form from the contexts in which it occurs.

The function of sub-word substance entities such as *-s* in the example *the cats eat some food* can also be deduced by means of contextual information and distribution throughout the language.[10] In the first place, the form *cats* is analysed as plural by virtue of its syntactic environment. Moreover, the suffix *-s* may be associated with plural functions according to the pattern of morphological variation and the distribution of fully-inflected wordforms in the paradigm of the lexeme {CAT}, and/or in the inflexional paradigms of other nouns. That is, once a speaker has encountered other forms which exhibit a similar implicational relationship (whereby [NOUN] can be associated with singular functions and [NOUN + s] can be associated with plural functions), they may extend this pattern to novel lexemes.

It must at this point be stressed that substance entities are *not* morphemes, though they may seem to be in some ways similar. A 'morpheme' is a minimal entity in language which *bears* meaning. That is, in the example offered above, the *-s* of the wordform *cats* would be analysed as encoding plurality. In a morphemic account (as discussed in chapter four), the language-user supposedly has this form stored along with rules about its distribution with relevant lexemes (here: count nouns). The combination of the lexical stem and the inflexional exponent *-s* 'add up' to the meaning 'multiple domesticated felines'; in this way, morphemes are compositional in terms of both form and meaning.

As a substance entity on the other hand, the sub-word form *-s* does not *encode* the function 'plural'. Instead, where it consistently occurs in a language in combination with particular word-types (e.g., count nouns) and in particular syntactic environments (for instance, where the verb agreement indicates plurality or where the pragmatic context indicates that a noun is to be analysed as plural), the speaker may abstract this suffix away from fully-inflected

[10] For brevity, I have omitted the phonological variants of the plural /s/. The distribution of the allomorphs /s/, /z/, and /ɪz/, however, is automatic in English, so even knowledge of the relationship between forms such as *dog ~ dogs* may aid a language-user in identifying and utilizing the patterned alternation associated with 'singular' vs. 'plural' functions.

wordforms, segment it from the fully-inflected wordform, and *associate* it with plurality. It is only with reference to and knowledge of other forms in the language which exhibit an alternation *-Ø ~ -s* in a context which differentiates number that this element can be associated with plural functions.

Utilizing substance entities, we may 'sidestep' the problems associated with identifying morphemes in language. That is, if there are two phonologically distinct sub-word substance entities in a language which can both be associated with the function 'plural', for instance, we do not have to 'derive' one from the other by means of complex *ad hoc* phonological processes. The Finnish partitive exponents, for instance, - *-A*, *-tA*, and, in the singular only, *-ttA* - cannot easily be derived from a single underlying morph which is selected by the grammatical or phonological environment in which it occurs. However, they can each be associated with 'partitive' functions and are transparently formally-related even in the modern language. Here, therefore, they are considered sub-word substance entities, despite a slight disturbance in the mapping of function and form.[11] Moreover, recognizing substance entities does not necessitate that all functions be expressed by invariant material since, as will be outlined in more detail in §5.1.2, all languages that are characterized by sub-word substance entities are also necessarily characterized also by schematic patterns.

The most important difference between 'morphemes' and 'substance entities' for our purposes is that the latter refers to a biuniqueness of form~function mapping at any and every level; the former refers only with sub-word (or monomorphemic) elements. That is, a substance entitity can be found at the sub-suffixal (the Finnish excessive; the novel Veps source case-forms), suffixal (the Finnish inessive exponent *-ssA*), word (Romanian inflected nominals), periphrastic, and even idiom levels (the latter two of which have not been

[11] Acquisition evidence for the status of the Finnish partitive allomorphs as sub-word substance entities will be offered in §5.2.1.

discussed in this study). I am here concerned largely with 'sub-word' and 'word-level' substance entities.

Schematic patterns are by their very nature often much more difficult to identify and 'list' in the same way as substance entities. Umlaut, ablaut, and consonant gradation all constitute examples of schematic patterns, since none of these instances of morphophonological variation can be easily segmented away from the speech stream in the same way as substance entities. For instance, the form *geese* in the English sentence given above is here analysed as a word-level substance entity, just like the form *cats*, since this form can be extricated from the speech stream and consistently associated with the meaning 'multiple web-footed birds slightly larger than ducks and smaller than swans'.[12] However, unlike the form *cats*, the form *geese* cannot be analysed as 'plural' by virtue of the presence of the word-final sibilant. Instead, the vowel sound /i:/, contrasting with the vowel sound /u:/ found in the singular form *goose*, discriminates number values in this lexeme. Thus, this form can only be interpreted as plural a) by reference to its function in the syntactic context, and/or b) by reference to the morphological paradigm and the abstract relationship between the inflected forms of this lexeme. The deduction of the singular form would require reference to the implicational relationships that exist between other lexemes such as {TOOTH}, where knowledge of the singular *teeth* and plural *tooth* may help a language-user to produce the correct singular *goose* from the plural *geese*. But the singular form of the lexeme {GOOSE} cannot be deduced by stripping away an overt substance entity (such as *-s* from the form *cats*); there is no way to 'deconstruct' the wordform *geese* in order to deduce the correct singular form.

In addition to these sorts of abstract alternations that exist between fully-inflected wordforms, I argue that the case/number suffixes in the Latin inflexional case system

[12] Adapted from the Oxford English Dictionary definition of 'goose': https://www.oed.com/view/Entry/80029?rskey=5oSKRO&result=1&isAdvanced=false#eid. Accessed 12.04.2019, 16:35.

constitute further examples of morphological variation which are better described as examples of schematic patterning. It has already been demonstrated in chapter four that a principled splitting of stem and suffix is not always possible, that there is very little consistent phonological similarity across declension (for instance, the genitive singular desinences *-ae*, *-ī*, *-is*, *-ūs*, and *-(e)ī* in the five major declensions), and that there are frequent instances of homophony (for instance, the element *-ī* can occur with forms encoding genitive singular functions as well as with forms encoding nominative plural functions). These case/number suffixes in Latin do not, therefore, constitute examples of substance entities, as the form~function mapping is disturbed in numerous ways. Intriguingly, they do provide more 'abstract' information, not about the function of a form, but about the shape of other forms in a lexeme's paradigm. This will be discussed further in §5.2 in reference to 'indexical' versus 'symbolic' information.

Despite the claim that these types of morphological alternation should be considered qualitatively different, it seems from the foregoing as though both substance entities and schematic patterns 'work' in similar ways, in particular that they are not informative in isolation, but rely on contextual information and/or knowledge of the paradigm to determine how the morphological variation may be exploited in the deduction of novel forms. The inflexion of previously unencountered wordforms in both of the English examples *cats* and *geese* (where plurality is encoded by means of the addition of a sub-word substance entity and where plurality is encoded by schematic sub-word formal variation) seems always to require reference to the syntactic context, to the paradigm, or to both. That is, it is not sufficient to say that the form *cats* is plural because it ends in a sibilant *-s*, since there are many other forms in English which end in a final sibilant and which are singular (e.g., *goose*). Syntactic information and/or reference to another form in the paradigm are required in order

to correctly interpret a form as singular or plural. Why, then, is it necessary to distinguish these two types of inflexional variation?

As I will demonstrate in the following sections, there are differences in the ways that speakers exploit substance entities and schematic patterns in language use which become clearer when the diachronic Italic and Finnic data are discussed in more detail. Once a language-user has met and recognized the distinct morphological structures found in a given language, they utilize them differently in the production of novel inflected forms. Thus, recognition and understanding of these distinct morphological structures offer insight into the developments of the inflexional case systems of the languages under consideration.

It is essential to note here that these different types of inflexional variation do not necessarily 'match' different morphological typological structures. For instance, typologically both Latin and Romanian may be described as 'fusional', and yet in Romanian, word-level substance entities which are stored and reproduced whole in language use are much more important than in Latin, where implicational patterns often provide sufficient information in the production of novel inflected forms in a way that abstract relations in Romanian do not (for instance, in the formation of the plural of a nominal). These differences are crucial to understanding why we find certain developments in the Italic and Finnic varieties discussed in this study.

5.1.1 SUBSTANCE ENTITIES, SCHEMATIC PATTERNS, AND REFERENCE TO THE PARADIGM

It is clear that reference to the paradigm is necessary where forms are characterized by schematic patterns, as has been discussed in the previous section and in §4.1.2. However, as has also been pointed out, even where form~function is direct, reference is needed to other inflected forms of a lexeme and to their distribution in the paradigm in order to generate novel forms correctly. This is demonstrated by the following example from Veps:

ekran ‘shield’

man ‘earth’

avoin ‘open’

sigan ‘pig’

Each of these forms ends in *-n*; generally, where this suffix occurs with the common stem of a lexeme, it can be analysed as the genitive singular exponent and is easily and consistently associated with ‘genitive singular’ functions. However, in certain lexemes, a final *-n* is found in the nominative singular form. From consideration of the surface forms offered above, there is no principled means of determining which of these forms encode nominative singular functions and which encode genitive singular functions without reference to other forms in the inflexional paradigm (or to syntactic clues). Knowledge of just one other inflected form in the paradigm permits a language-user to unambiguously determine the functions encoded by the forms offered above:

nominative singular *ekran*, e.g., inessive plural *ekranis*

e.g., partitive plural *maid*, genitive singular *man*

nominative singular *avoin*, e.g., ablative singular *avoinudelpäi*

e.g., comitative singular *siganke*, genitive singular *sigan*

The genitive singular form *man* may initially be analysed as nominative singular, but knowledge of the partitive plural form *maid* provides sufficient information to falsify this assumption, as a language-user with knowledge of other partitive and/or plural forms will be able to decompose *maid* into a lexical stem, a plural marker *-i-* and a partitive exponent *-d*, thus leaving the stem *ma-* which is also found in the form *ma-n*. If the final *-n* of this form

were found in the nominative singular, it would also be present in the stem of the plural (plural stem ***man-i-*, for instance). Since it is absent in the form *maid*, the form *man* may be analysed as the genitive singular form.

In a language characterized by sub-word substance entities, a language-user must still make reference to other forms in the paradigm in order to correctly deduce the function or form of a novel nominal. One may well ask: how is this any different from the necessity of referring to the paradigm in the utilization of schematic patterns in novel form production?

The importance of a distinction between substance entities and schematic patterns is clearer in the *production* of novel inflected forms than in the *interpretation* of inflected forms. In the first place, inflected forms in language are extremely rarely encountered without context, which means that an audience (either one listening to speech or reading a text, for instance) will have other syntactic clues which will aid in the deduction of the function of a novel form.

the <u>cats</u> eat the food

the <u>geese</u> eat the food

Even without any other information about the formation of plural nouns in English, an audience can tell that these are plural because they occur with the third plural form of the verb {BE}. It does not matter that the relationship between the singular *cat* and the plural *cats* is that of the presence or absence of a final sibilant, or that the relationship between the singular *goose* and the plural *geese* is that of the alternation of the vowel: here, these forms can be correctly interpreted as plural by virtue of the context in which they occur. Where a speaker knows the plural form and the hearer comes across it in context, therefore, the difference between substance entities and schematic patterns does not appear to be

significant. The difference between schematic patterns and substance entities is most important when it comes to consideration of the Paradigm Cell-Filling Problem (§4.1.2), as well as understanding the role played by the competing pressures of predictability and discriminability in language use.

At a sub-word level, substance entities provide more certainty in the production of previously unencountered inflected wordforms than schematic patterns. In Veps, for instance, in the paradigm of a lexeme with an invariant lexical stem such as *hebo-* 'horse', the nominative singular form and the minimally-entropic inflexional case and number markers permit a language-user to deduce all forms in the paradigm with certainty. These forms are not as highly discriminated as the suffixal exponents found across declension in Latin, for instance, but they do exhibit a great deal of predictability. Communicative information is thus sacrificed for increased predictability.[13]

In Romanian, on the other hand, knowledge of one fully-inflected form does not provide such maximally-predictive information about the deduction of the second (or, very occasionally, third) form in the inflexional paradigm of a feminine nominal. The genitive-dative singular form (and, thus, the plural form), for instance, can only be produced 'correctly' where it is known, as the abstract relations that exist between other word-level substance entities do not provide maximally-predictive information about the other form(s) in the paradigm of feminine lexemes. However, the fully-inflected wordforms in the Romanian case system are much more highly discriminated from one another than the forms in the Veps nominal paradigm, often exhibiting differences in both stem and suffix, which makes them more easily distinguished by the hearer. This increased communicative information is

[13] There are, of course, other considerations in communication which provide 'predictive' information. For instance, if two strangers meet at a bus station, the common physical location and intent (of catching a bus) may reduce the uncertainty of the subject of a conversation that may follow. I am concerned in this study only with the predictive information which reduces uncertainty about the production of novel inflected forms.

accompanied by a concomitant sacrifice in predictability. Substance entities thus provide instances of sub-word regularity which can aid in the production of novel inflected forms, whilst schematic patterns may be more highly discriminated (although not always, as will be seen in §6.2 in reference to old Gallo-Romance), but are not necessarily always as useful in the deduction of previously-unencountered inflected wordforms.

Finally, there is an important difference between the implicational patterns which exist between two or more fully-inflected wordforms and the distribution of forms within the inflexional paradigm, the latter of which I call 'paradigmatic schemata'. The necessity of distinguishing these two can be seen most clearly in the case system of Romanian. Here, implicational patterns that exist between inflected forms provide some information about the shapes for novel forms, but it is never minimally-entropic in feminine nominals. For instance, on the model of *casă ~ case* 'house' we might suggest that the nominative-accusative singular form of a nonce lexeme †*fasă* might be †*fase*, whilst on the model of the implicational pattern exhibited by forms such as *fată ~ fete*, the form produced is instead †*fese*. The uncertainty regarding the inflexional behaviour of a novel form in Romanian is neatly illustrated by loanwords, where French *panne* 'breakdown' has been borrowed into Romanian as *pană ~ pene* 'breakdown (of a motor)' on the pattern of native *pană ~ pene* 'feather', but the loanword *masă ~ mase* 'mass' (from French *masse* 'mass') does not follow the same inflexional pattern as native *masă ~ mese* 'table'.[14] There therefore appears to be vacillation around the selection of the patterns in the inflexion of novel lexemes.

The distribution of these forms in the paradigm, by contrast, is very regular in the language. Knowing, for instance, a genitive-dative singular form *cozi* 'tail', and having an awareness of the morphomic distribution of forms in Romanian feminine nominals, a language-user may correctly determine that the form of the plural is the same, *cozi*. It thus

[14] My thanks to Martin Maiden (p.c.) for these examples.

appears as though schematic patterns and paradigmatic schemata may be recognized and exploited by language-users in the deduction of the correct inflected form of a lexeme, but they provide slightly different information: schematic patterns are more important for the deduction of the phonological shape of an inflected lexeme, whilst paradigmatic schemata are more important in the distribution of different (and the same) forms within a paradigm.

5.2 SYMBOLIC AND INDEXICAL INFORMATION

In addition to different types of morphological variation found in a language, there are also different *types of information* which can be encoded by substance entities and schematic patterns. In the first place, substance entities, it has already been explained, are those sub-word units which can be reliably associated usually with just one feature value, or one consistent set of feature values. So, for instance, in Finnish, the suffix *-llA* is associated with 'adessive' functions; the exponent *-n* is associated both with 'genitive' and with 'singular' functions. However, these exponents provide absolutely no information about the inflexion class to which the nominal belongs, since they are found across all declensions. In contrast, the inflexional variation in Latin frequently provides a language-user with information with which the Paradigm Cell-Filling Problem may frequently be solved. Formal variation in the suffix therefore provides different types of information, whether about the *function* of a form or the *inflexional behaviour* of a given lexeme. In order to differentiate these different types of information, I will adapt and employ two terms introduced by Charles Sanders Peirce as part of his semiotic theory, namely: 'symbol' and 'index'. I will first outline how these terms are defined by Peirce and then explain how they will be used here.

In Peirce's theory of semiotics, there are three types of signs: icons, symbols, and indices. Icons represent or imitate referents; in language, onomatopoeias may be considered icons. These will not be considered further here. Symbols are signs in the Saussurean sense, and can

be used to describe words which stand in an arbitrary relation to the meaning(s) they encode. For instance, nothing about the phonological form *cat* 'imitates' its referent in the same way as an icon (compare the relationship between onomatopoeic *bang* and the sound it is intended to represent). Finally, indices in Peirce's semiotic theory "show something about things, on account of their being physically connected with them."[15] The most common example of an index is that of smoke signalling fire. These two things are connected implicatively: smoke does not 'imitate' fire, nor stand in a completely arbitrary relation to it.

The terms 'symbolic' and 'indexical' are used here to describe the different types of information signalled by certain morphological forms. That is: substance entities are considered to be 'symbolic', whilst schematic patterns are more frequently 'indexical' (although not always, as will be discussed below).

In the first place, wherever a Veps nominal ends in *-l* (in combination with the common stem), the suffix signals that the form is to be interpreted as encoding 'adessive' functions, regardless of inflexion class, number, lexeme, or word class. This form therefore stands in a direct, albeit arbitrary, relation to the functions it signals. In this way, it exhibits the same 'symbolic' relation of form and function (or Saussurean signifier~signified relation) as *cat* and its referent. Thus, any units of analysis above, below, and at the level of the word, which exhibit no imitative relationship to the function(s) of the signal, are considered here to constitute examples of 'symbols'.

The suffixal forms found in Latin, on the other hand, cannot be as easily segmented away from fully-inflected wordforms and associated with particular functions. In this way, they do not appear to constitute 'symbols' in the same way as a Veps exponent *-l*. However, Latin suffixal alternations are generally more informative than those found in Veps about other

[15] Peirce (1998:5).

considerations, including number, inflexion class, and, sometimes, gender. The schematic patterns exhibited by the formal variation in the suffix of Latin nominals may be described as 'indexical'. That is, a suffix *-ī* in Latin does not provide unambiguous information about the case/number values encoded by a fully-inflected wordform. It does, however, provide other information, particularly where it can be associated with the functions it encodes. The Veps adessive *-l*, even in combination with knowledge about the functions it encodes, does not provide any information about inflexion class, for instance, because of its generality, whilst in Latin the combination of the form in *-ī* and its encoding nominative plural functions indicates unambiguously that the lexeme belongs to the second declension (and also that it is very likely to be masculine and cannot be neuter). In this way, the nominative plural exponent *-ī* can be associated with the inflexional pattern exhibited by second declension lexemes, just as a smoke is associated with fire: the exponent does not *imitate* its function (as an icon would), nor unambiguously relate form and meaning (the same element is found in the genitive singular form of second declension lexemes); instead, it provides implicational information about the shape of other forms in the paradigm. Many of the schematic patterns found in the Latin case system are therefore analysed as providing indexical information that may be exploited in solving the Paradigm Cell-Filling Problem.

Intriguingly, there appears also to be correlation between symbolic and indexical information, and the competing communicative pressures of predictability and discriminability. Where a form at the sub-word level stands in a biunique (albeit arbitrary) relationship with a function, it provides more predictability that can be utilized in the production of novel forms. A more highly discriminated form, such as those found in the Latin case system, however, also often appears to provide more indexical information.

It will be suggested in chapter six that case systems which tend to lose morphological case distinctions will be characterized by *neither* sub-word substance entities *nor* a high degree of

indexicality in its inflexional variation. For instance, sub-word substance entities are not found in the inflexional case system of old French and old Occitan, and, whilst there is evidence of an implicational pattern which was recognized and exploited by speakers (i.e., the chiastic distribution of *-s/-Ø* in the masculine nominals - §2.2.1), there are no instances in which the shape of the suffix, in addition to the function it signals, unambiguously provides information about the inflexion class, number, or gender of a lexeme. The morphological expression of case has been completely lost from nominals in both these varieties.

Similarly, although stable at an earlier stage of its life, the Latin case system lost crucial morphological distinctions in nominals as the result of regular phonological changes, which caused a concomitant reduction in the indexical information signalled by suffixal variation. Since there were no sub-word substance entities that could be utilized in 'constructing' novel inflected forms, and since the implicational relations signalled by indexically-informative suffixes were lost as the result of phonetic erosion, there was no longer near minimally-entropic information on which to rely in language-use. Invariance at the word-level offered more certainty for language-users (although distinctions were usually maintained across number), and these could be stored as symbols which arbitrarily, but minimally-entropically, related form and function. This will be discussed further in chapter six.

5.3.0 EVIDENCE FOR SUBSTANCE ENTITIES

In the following section, I will present data which support my contention that speakers recognize and exploit sub-word substance entities in the production of novel forms where there is transparent sub-word structure in the grammar. I will also support the suggestion that substance entities behave differently from schematic patterns and paradigmatic schemata in diachrony. I focus on evidence from Finnish, since, of the Finnic languages discussed here, it

is the best researched in terms of L1 and L2 acquisition,[16] and is also characterized by an abundance of sub-word substance entities.[17]

Many psycholinguistic tests have been conducted in order to investigate how inflected forms are interpreted and produced - whether by the decomposing of fully-inflected forms into meaningful sub-word elements (or, in production, the concatenation of these elements), by utilizing internalized 'rules', or by extending inflexional patterns analogically - and how they are produced. Although the conclusions offered on this subject are not always clear (or may be interpreted in a number of ways), many psycholinguistic experiments provide evidence to support the suggestion that speakers can recognize and make use of sub-word recurrent partials in the production of novel forms, even in languages with an abundance of fusional structure such as English (e.g., the final sibilant of plural nouns). The results of many experiments suggest that sub-word entities are identified by and exploited by language-users (e.g., Taft 1979, Feldman 1991, and Laine et al. 1995), but that substance entities are not *required* in the production and/or interpretation of novel inflected forms.

5.3.1 EVIDENCE FROM L1 LEARNERS OF FINNISH

Evidence from L1 learners of Finnish indicates that sub-word substance entities, such as case suffixes, are recognized, stored, and used constructively in the production of many novel forms. The nominative singular is generally the first form acquired and used by an L1 learner, followed by the grammatical cases (genitive and partitive), and then the illative, all in the singular.[18] The first of these forms are generally learned and stored. When children begin to create inflected forms for themselves, several different types of errors occur when

[16] The acquisition of Estonian has also been extensively studied by, e.g., Argus (2007, 2009, and 2012).
[17] Although L2 data are not always considered in investigations about the production of inflected forms, here I include such evidence here to demonstrate that both L1 and L2 learners recognize and utilize the abundance of substance entities in the production of novel forms in Finnish, and that the tendency towards construction is thus not just a reflexion of the 'agglutinative preference' of child language.
[18] Laalo (2009:54).

morphological variation is found. Errors include: 1) lack of consonant gradation, 2) use of the incorrect stem shape, and 3) back formation of nominative singular from an inflected form (i.e., the deduction of a non-existent stem shape). Finally, although less frequent than errors in stem shape, where a case suffix exhibits formal variation across number and/or declension, there are instances in which children 'select' the incorrect inflexional exponent. Where there is invariance below the level of the word, however, morphological errors are far less frequent.

First, some children form the genitive singular with the suffix *-n* in addition to the strong stem of a nominal, where the weak stem would be expected: ***ukkon* for *ukon* (nominative singular *ukko* 'old man') and ***tyttön* for *tytön* (nominative singular *tyttö* 'girl').[19] Other children form the genitive singular without the final *-n*, where, in certain nominal types, the nominative and genitive singular are distinguished by consonant gradation alone: ***lapu* for *lapu-n* (*lappu* 'bib') and ***meko* for *meko-n* (*mekko* 'dress'). In early acquisition, therefore, final *-n* and the stem change (here, marked by consonant gradation), jointly or severally, seem to be sufficient to discriminate the genitive from the nominative (as well as the other early case-forms, partitive *-(t)A* and illative *-Vn*, which are both formed on the strong stem for the lexemes given above). It therefore does not seem as though there is any evidence that children treat the suffix *-n* as 'encoding' the genitive since the stem changes are also perceived by Finnish children as in some way 'meaningful'. However, the tendency to form the genitive singular with just the weak stem tends to happen early (~ 1 year 7 months) and is 'corrected' fairly swiftly (by ~ 1 year 10 months), when children may continue to make mistakes in the formation of the stem shape, but the final *-n* is found more and more consistently.

There are a number of other stem alternations found in Finnish nominals which do not have to do with consonant gradation (described in §3.1.1). For instance, the *i/e*-stem nominals

[19] This reflects the 'agglutination' preference found in child language. Dressler (2003:10).

and *s*-stem nominals cause problems for Finnish children. Young children often give genitive singular forms such as ***käsin* for the lexeme {KÄSI} 'hand', the nominative singular form of which is *käsi*. This lexeme in adult language follows the inflexional pattern of a *vesi*-nominal, but in child language, instead of the expected genitive singular form *käden*, we find strict agglutinative morphological structure.[20]

Consonant stems in general "have been observed to be difficult for Finnish children".[21] Niemi and Niemi report some erroneous partitive singular forms for consonant-stem nominals: ***ukkonena* and ***ukkonesta* for *ukkosta* (nominative singular *ukkonen* 'thunder'); ***keltanea* for *keltaista* (nominative singular *keltainen* 'yellow'); and ***pieneä* for *pientä* (nominative singular *pieni* 'small').[22] Here there is a preference for simple agglutination of a partitive singular exponent *-A* and the nominative singular stem (i.e., *ukkonen* + *a*, *pieni* + *ä* = *pien<u>e</u>ä*), or for a stem minus a final *-n* (*kelta(i)ne* + *a*), which here might have been analysed as the genitive singular exponent by the child.

Similarly, there are some examples of the erroneous back-formation of nominative singular forms (i.e., the deduction of a nominative stem from a common stem) from, e.g., the genitive or partitive singular forms: ***tiile* for *tiili* 'brick, tile' (compare genitive singular *tiile-n*); ***eteise* for *eteinen* 'hallway' (compare genitive singular *eteise-n*); and ***avaime* for *avain* 'key' (compare genitive singular *avaime-n*).[23] In these examples, the children have presumably met other inflected forms of these lexemes, such as the genitive singular, and have attempted to deduce the nominative singular form. This requires recognition and segmentation of the invariant inflexional suffixes as well as recognition of the pattern whereby nominative singular forms are characterized by a *lack* of overt suffixal marker. For

[20] Laalo (2009:61).
[21] Niemi and Niemi (1987:66). Also Nesser (1981:59) and Karlsson (1982:199).
[22] 1987:66.
[23] Niemi and Niemi (1987:64).

some nominals, indeed, removing the final inflexional elements of a singular inflected form does produce the correct nominative singular form: genitive singular *nalle-n* 'teddy' ~ nominative singular *nalle*; genitive singular *omena-n* 'apple' ~ nominative singular *omena*. In the examples given at the beginning of this paragraph, however, simply removing the inflexional suffix does not produce the correct nominative singular form. Nevertheless, the pattern 'nominative singular = *-Ø*, all other forms = *-cx*', appears to be salient even very early in L1 acquisition.

Finally, although the majority of case suffixes in Finnish do not exhibit formal variation across number and/or declension, two of the four case-forms first acquired by L1 learners can be associated with three distinct exponents:

Nominal type	**Partitive singular**	**Illative singular**
omena 'apple'	*omena-a*	*omena-an*
maa 'land, earth'	*maa-ta*	*maa-han*
perhe 'family'	*perhe-ttä*	*perhee-seen*

Table 5.1: Finnish partitive and illative singular exponents

In the early acquisition of Finnish, forms offered by children exhibit fluctuations in the selection of the partitive singular case desinence,[24] although even from an early age, *-A* seems to be preferred in the production of novel forms. For instance, "[i]n *kastike ... a* (age 4.1) where ... indicates a pause, the child himself is unsure how to inflect *kastike* 'sauce' in the partitive singular (the correct form would be *kastiket-ta*, stems: *kastikkee-* V, *kastiket-* C). This uncertainty is reflected in the long pause at the boundary of the stem and suffix."[25] The child in this study selects the partitive form *-A* with an unfamiliar lexeme (where knowing only the nominative singular form *kastike* offers no indication as to whether it follows *nalle-* or *hame*-type nominals). It may be that the form **kastikea* is an overgeneralization of the

[24] Niemi and Niemi (1987:72f.).
[25] Niemi and Niemi (1987:66f.).

nalle-type nominal in young speakers' language.[26] The addition of the marker *-A*, following a pause, to the (incorrect) lexical stem in order to produce the partitive singular form suggests that children can and do to a certain extent *construct* novel forms from sub-word entities, particularly where these can be consistently associated with particular functions.

Despite the fluctuations in the distribution of these case exponents, there is still evidence that these three forms, *-A*, *-tA*, and *-ttA*, are each closely associated with 'partitive' functions. The functions of the partitive are mastered relatively young,[27] and, although there are errors in the formation of the partitive singular, it appears as though these three desinences are closely related to partitive functions in the mind of L1 learners. For this reason, I consider them sub-word substance entities, despite exhibiting formal variation, since they can all be easily be segmented away from a fully-inflected wordform and associated with partitive (singular) functions. A similar conclusion may be reached for the illative singular, where the illative singular form *-Vn* tends to generalize over the other forms *-hVn* and *-seen*.[28]

In all instances where nominals have been inflected 'incorrectly', the errors mostly have to do with the shape of the stem (where formal variation within and across declensions is found). The genitive singular forms given here, for instance, (almost) all exhibit a final *-n* regardless of the alternations that are found in the lexical stem. Although the shape of the stem presents some difficulties for L1 learners, therefore, they recognize very quickly the directness of the mapping of genitive singular function and the suffix *-n* and rely on it in the production of novel forms. Similar observations also hold for the other early case-forms, the partitive and illative in the singular.

[26] Räisänen (1975), Itkonen (1977), Nesser (1981), Karlsson (1982), Niemi and Niemi (1987).
[27] Laalo (2009:72).
[28] Niemi and Niemi (1987:68f.).

5.3.2 EVIDENCE FROM L2 LEARNERS AND NATIVE SPEAKERS OF FINNISH

According to evidence gathered by Martin (1995) and Hakanen (2006), L2 learners of Finnish frequently make similar mistakes - at least, in writing - to those found in children's speech when it comes to inflecting nominals.[29] The two studies focused on here present evidence about the production of genitive singular, genitive plural, partitive singular, and partitive plural forms from a given nominative singular form (both real and nonce forms were provided). The partitive plural forms and genitive plural forms are generally the most difficult to form given knowledge of the nominative singular form of a lexeme alone.

There are a number of differences in the responses offered by L2 compared with L1 language-users across the data sets:

- a greater proportion of native Finns tended to select the 'expected' inflected forms for both real and nonce words than the L2 learners;
- the L2 learners tended to offer a greater variety of answers than the native Finns;
- the native Finns' mistakes tended to consist of: unexpected stem shapes; dialectical alternants; and plural forms for singular (or vice versa for non-count nouns);
- the L2 learners' mistakes tended to consist of: unexpected stem shapes, and incorrect vowel harmony.

However, there are also important similarities between the two groups in each data set. Despite differences in the stem shape or suffixal vowel, for both L1 and L2 speakers:

- the genitive singular forms nearly always ended in *-n*;

[29] The drawbacks of referring to written examples are clear. In particular, it is likely that, given time to reflect on their responses, both L1 and L2 learners had more time to come up with the expected inflected form; in speech, they may have made more frequent, or at least different types of, errors. Nevertheless, the written data is more readily available and easier to interrogate, and thus serves as the basis of the discussion here. Anecdotally, from my own experience, the responses offered in writing reflect the same general tendency as is found in the spoken language: stem errors are a great deal more frequent than errors in the suffix.

- the genitive plural forms overwhelmingly ended in *-ien, -ten,* and *-den*;
- the partitive singular and plural forms almost all ended in *-(t)A*.[30]

What is clear from the data offered by Martin (1995) and Hakanen (2006) is that there is a relative stability of the case suffixes when compared to the shape of the lexical stem with which they occur. However, there is some vacillation in the selection of the 'correct' case exponent where a value is characterized by distinct forms across number and/or declension. In Martin's (1995) data, the genitive singular morph *-n* is invariant, and this is almost always captured by language-users. The partitive in the plural, on the other hand, can have two different shapes: *-tA* and *-A* (immediately preceded by the plural form *-i-*, orthographically *-j-* between vowels). Overwhelmingly, the partitive form *-A* is most common in both nonce and real words for L1 (adult) and L2 speakers:

	L2		**L1**	
	Nonce words	**Real words**	**Nonce words**	**Real words**
-A	80%	78%	87%	83%
-tA	15%	16%	12%	17%
Percentage correct forms given	35%	44%	73%	95%

Table 5.2: Distribution of partitive exponents in Martin (1995)'s data[31]

The higher percentage occurrence of the desinence *-A* in the formation of the partitive plural in nonce words for both L1 and L2 language-users may reflect the high percentage of the occurrence of partitive *-A* (as opposed to *-tA*) in real words in Finnish (83% of Martin's words are expected to end in final *-A*, 17% in *-tA*). Although speakers (both native and non-

[30] The data set presented by Hakanen (2006) does not include any lexemes which should take *-ttA* in the singular.

[31] Only entirely correct forms (i.e., with correct stem shape and case suffix allomorph) have been included in the 'correct forms given' row. Here, errors in vowel harmony - largely found in L2 learners' responses - have not been treated as errors, since the alternation appears to be automatic in speech, even where it is not captured orthographically; null responses, forms in *-ttA*, and non-partitive forms have been omitted from consideration, which is why the numbers in table 5.2 do not always add up to 100%.

native) appear to be aware that there are two exponents for the partitive plural, they may also be sensitive to frequency considerations.

In Hakanen's (2006) data, similar observations can be made about the distribution of the genitive plural desinences.[32] In Finnish, the genitive plural has up to five different suffixal shapes: *-tten*, *-ten*, *-den*, *-ien*, and *-in*.[33] The majority of nominals have more than one genitive plural form. Hakanen (2006) looks specifically at the formation of the inflected forms of *s*-stem nominals, which have the following genitive plural exponents:

	-tten	***-ten***	***-den***	***-ien***	**Gloss**
Vs : VV	*rukiitten*	*X*	*rukiiden*	*X*	rye
s : ks	*X*	*tungosten*	*X*	*tungoksien*	crowd
(U)Us : (U)Ude	*X*	*X*[34]	*X*	*raukkauksien*	love

Table 5.3: Genitive plural exponents in Finnish *s*-stem nominals

In the formation of real and nonce words, the use of one of these exponents is almost universal (although some desinences are more frequently used than others, as will be discussed below). The locus of greater inflexional error is that of the stem, where nonce forms which are 'expected' to behave like *Vs* : *VV*-type nominals are frequently inflected as though they belong to the *Vs : ks* class (the larger and more productive declension class). However, since there are a greater number of exponents which can be associated with 'genitive plural' functions than there are which can be associated with 'partitive' functions, there is some vacillation in the selection of the particular 'genitive plural' desinences across lexemes and L1 vs. L2 speakers, as shown in table 5.4:

[32] Conclusions drawn from the forms considered here do not necessarily reflect patterns that might occur in other (non-*s*-stem) nominal types.
[33] None of the nominal types considered here exhibit any forms in *-in*. Some respondents have given a form in *-in*, but these are very infrequent. They are then included in the 'other' count in table 5.4.
[34] Although grammars do not give a form ***rakkausten*, other nominals of this type can have an alternative form in *-ten*, e.g., *avaus ~ avauksien ~ avausten* 'opening'. (Ulla-Maija Forsberg, p.c.).

		L2		L1	
		Nonce words	**Real words**	**Nonce words**	**Real words**
Vs : VV	***-tten***	4%	6%	0%	2%
	-ten	7%	4%	15%	9%
	-den	17%	18%	33%	81%
	-ien	**35%**	**39%**	**46%**	4%
	OTHER	37%	33%	6%	4%
	Percentage correct/expected forms given	2%	15%	17%	84%
s : ks	***-tten***	7%	5%	0%	0%
	-ten	5%	6%	30%	36%
	-den	8%	6%	4%	5%
	-ien	**44%**	**50%**	**56%**	**57%**
	OTHER	36%	33%	10%	2%
	Percentage correct forms given	17%	28%	83%	90%
(U)Us : (U)Ude	***-tten***	4%	4%	0%	0%
	-ten	4%	6%	4%	2%
	-den	3%	7%	28%	10%
	-ien	**52%**	**51%**	**61%**	**84%**
	OTHER	37%	32%	7%	4%
	Percentage correct forms given	23%	33%	48%	83%[35]

Table 5.4: The distribution of Finnish genitive plural exponents

Highlighted in table 5.4 are the desinences that are expected with particular nominal types. Although the genitive plural exponents *-tten*, *-ten*, *-den*, and *-ien* all occur in the data set considered, they do not occur with equal frequency. For both real *and* nonce words, and for all inflexion classes, L1 and L2 speakers appear to prefer the form *-ien*. (The only exception is the real words of the *Vs* : *VV* type for L1 speakers, where it is likely that whole forms will have been previously encountered and reproduced whole.) Thus, in the two classes where *-ien* is one of two optional genitive plural exponents, it is preferred (over *-tten* and *-ten*), and

[35] Here, as in table 5.2 above, a form is only included in this percentage if the entire form is correct. Where there are only errors in vowel harmony, these forms are included amongst the correctly inflected forms.

where it is not expected to occur (in the *Vs* : *VV* class) it is only less preferred where fully-inflected wordforms are likely to have already been encountered by L2 speakers. Intriguingly, the exponent *-ien* is clearer in its syntagmatic structure than *-tten, -ten*, and *-den*, as it transparently consists of a plural marker *-i-* and a genitive marker *-n* (with *-e-* intervening).[36] It is possibly partly for this reason that this desinence is generalizing, particularly in previously-unencountered forms, at the expense of the more opaque forms *-tten* and *-ten* (although *-den* is still relatively frequent, particularly in real lexemes belonging to the *Vs* : *VV* and *(U)Us* : *(U)Ude* classes (mainly for L2 speakers).

The most important observation to draw from evidence presented here is that the case suffixes appear to be relatively phonologically stable when compared to the stems with which they occur. It appears as though native speakers and L2 learners exploit the relative certainty exhibited by the direct mapping between the 'genitive singular' functions and the form *-n*, as well as the 'partitive singular' functions and the suffixes *-A* and *-tA*, even where the same certainty cannot be relied upon in the deduction of the lexical stem shape. Where a suffix is characterized by phonological variation, there tends to be a preference for one form over the other which, in the genitive plural, happens to be that form that can be most easily decomposed into an element *-i-* discriminating 'plural' functions and an element *-(e)n* discriminating 'genitive' functions.

'Uncertainty' in the inflexion of novel forms does not actually come from the lexical stem, but from the 'link' between the stem and the inflexional material that follows. For instance, in *s*-stem nominals, most of the lexical stem is invariant through the paradigm:

[36] The only other exponent which is more syntagmatically transparent is *-in*. Forms ending in this genitive plural exponent sound archaic, however, and the spread of this desinence may be inhibited by its stylistic flavour.

	***ajatus* 'thought'**		***vieras* 'guest, stranger'**	
	Singular	**Plural**	**Singular**	**Plural**
Nominative	*ajatus*	*ajatukset*	*vieras*	*vieraat*
Genitive	*ajatuksen*	*ajatuksien/ajatusten*	*vieraan*	*vieraiden/vieraitten*
Essive	*ajatuksena*	*ajatuksina*	*vieraana*	*vieraina*
Ablative	*ajatukselta*	*ajatuksilta*	*vieraalta*	*vierailta*

Table 5.5: Uncertainly in the 'link' between lexical stem and inflexional suffix

Across these two paradigms, apart from in the genitive plural cells, the inflexional material is identical; within each of the declensions, the forms *ajatu-* and *viera-* respectively remain phonologically similar. Instead, the sub-word 'part' which is not always predictable in the production of novel inflected forms is the final *-s* found in the nominative singular form and whether it alternates with *-ks-* or *-V-* (that is, a copy of the vowel which precedes it in the singular (and nominative plural), or *-i-* elsewhere in the plural paradigm). This uncertainty is reflected in the data collected particularly by Hakanen (2006): the case (and number) exponents provide an element of certainty in the inflexional case system; the 'link' between the lexical stem and the inflexional material is a locus of greater uncertainty.

5.4 EVIDENCE FOR SCHEMATIC PATTERNS

The previous section considered evidence from Finnish which indicated that sub-word substance entities can be - and frequently are - segmented away from fully-inflected wordforms and exploited in the production of novel inflected forms. Where the mapping of function and form is not direct, however, it is argued here that speakers do not - and, in fact, *cannot* - simply concatenate sub-word substance entities onto lexical stems in order to produce novel forms. Instead, alternative means, such as the implicational schematic patterns that hold between two or more inflected forms, must be used in the deduction of previously-unencountered items.

Evidence that language-users are sensitive to and utilize schematic patterns in the production of novel forms may be found in the inflexional behaviour of certain old Occitan

and old French nominals. In many ways, the morphological patterns exhibited in these old Gallo-Romance varieties appear to be very weak, since they are no longer found in the modern languages. However, the fluctuation of *-s/-Ø* that distinguishes number and/or case feature values appears to have been recognized by speakers of old French and old Occitan, at least in a sub-section of the system, as we find that the chiastic distribution of suffixal *-s/-Ø* which originally developed from regular phonological changes was extended analogically to *pere/paire*-type nominals in the singular as a means of distinguishing nominative from oblique:

	Etymological *paire* 'father'		**Old Occitan *paire*-declension**	
	Singular	**Plural**	**Singular**	**Plural**
Nominative	*paire*	**paires*	*paire(s)*	*paire*
Oblique	*paire*	*paires*	*paire*	*paires*

5.6: Inflexional patterns exhibited by old Occitan {PAIRE}

Through time, final *-s* came to be associated with nominative singular functions, particularly in masculine lexemes,[37] which suggests that speakers were sufficiently aware of the difference in form exhibited by the larger and more productive *murs*-class to extend the implicational pattern to other lexemes in the language which, historically, did not inherit the same formal alternation.

The development of the inflexional case system of Latin also provides some evidence to suggest that language-users recognize and utilize schematic patterns in the production of previously unencountered forms. The case system of the Classical language appears to have been very 'cohesive'. That is, frequently only two (or even one) inflected forms of a lexeme (in addition to the functions they encode) provide sufficient information for a language-user to determine the declension class to which it belongs. However, in those instances where a

[37] Discussed by Jensen (1976:124f.) in reference to old Occitan and by Pope (1952:313) in reference to old French; also outlined in §2.2.3.

given inflected form does not provide sufficient information about another form in the lexeme's paradigm, we find, as discussed in §2.1.3, the shifting of nominals from one inflexion class to another. In particular, the nominative singular and accusative singular forms of second and fourth declension lexemes exhibit the same suffixal alternation *-us* ~ *-um*. As a result, there are occasional errors in inscriptional evidence wherein fourth declension nominals inflect like second declension lexemes; moreover, in the modern Romance languages, the vast majority of etymologically fourth declension nominals now exhibit inflexional behaviour identical to that of lexemes which historically belonged to the second declension in Latin.[38] This transfer of Latin lexemes from one declension to another through time thus supports the supposition that inflexional patterns can be recognized by language-users and utilized in the inflexion of novel items.

5.5 EVIDENCE FOR PARADIGMATIC SCHEMATA

Evidence for the diachronic stability and importance of paradigmatic schemata comes from the Romanian case system. The diachronic stability of the morphomic pattern found in nearly all Romanian feminine nominals has already been noted in §2.3.3. The distribution of forms within the feminine paradigm is very regular across the language:

	Singular	**Plural**
Nominative-accusative	FORM1	FORM2
Genitive-dative	FORM2	FORM2

Table 5.7: Morphomic distribution of forms in Romanian feminine paradigms

For the vast majority of feminine nouns and adjectives, the distribution of the forms is absolutely predictable. Where a language-user has met and stored the two forms found in the lexeme's paradigm, and where they either know the functions encoded by these forms (or can infer the functions they encode by reference to the abstract relationship between the two

[38] As mentioned in §2.1.3, there are very rare instances of fossilized relics of the fourth declension in certain dialects of Italian.

fully-inflected forms), they can determine for certain the distribution of the different forms within the paradigm. Furthermore, there is historical evidence to suggest that this pattern is recognized and utilized by language-users, since any developments which affect, for instance, the form of the genitive-dative singular, will equally affect the forms in the plural, and vice versa.[39] Apart from the handful of exceptions (i.e., certain three-form mass nouns; invariant feminine forms), therefore, this example of a paradigmatic schema appears to be very powerful in Romanian and can be exploited by speakers in the prediction of certain forms. That is: where a speaker knows a genitive-dative singular form, they can with near certainty extend that form to the nominative-accusative plural and genitive-dative plural cells of the same paradigm. They also know that the nominative-accusative singular form is likely to be distinct.

Latin may also provide some evidence for the reality of paradigmatic schemata, though perhaps of a slightly different type from that found in Romanian. It was suggested in chapter two that the extension of the accusative plural form of a first declension nominal to the nominative plural cell may be an analogical development whereby the pattern of the distribution of third declension nominative/accusative plural forms has been extended to first declension nominals; this may have been facilitated by the phonological similarity of the endings *-ās* and *-ēs*. Here, the patterned distribution of forms in the paradigm is not analogically extended to other lexemes according to a similarity in gender value (as in Romanian), but instead because of similarity in the phonological structure of the suffix occurring with forms encoding 'accusative plural' functions.

This paradigmatic schema from Latin does not appear to be quite as powerful as the distribution of forms found in Romanian, since it does not tend to 'attract' lexemes from the second declension (even though we find a similarly structured suffix in these accusative

[39] Maiden (2015:43).

plural forms *-ōs* - §2.1.3). Nevertheless, consideration of this development appears to demonstrate that language-users recognized the pattern exhibited by third declension nominals and extended it to first declension lexemes.

5.6 CLOSING REMARKS

This chapter has introduced the terms ‘substance entities’, ‘schematic patterns’, and ‘paradigmatic schemata’, has explained how they will be used in the chapters that follow, and has attempted to provide from acquisition and comparative historical evidence for their behaviour in diachrony. It has also explained what is meant by ‘symbolic’ and ‘indexical’ information which can be encoded by distinct types of morphological variation.

It will be argued in the next chapter that proper understanding of the differences between substance entities, schematic patterns, and paradigmatic schemata, as well as the ways in which they are exploited by language-users in the production of inflected forms is essential to understanding the particular developments of the Italic and Finnic case systems. It will be shown, moreover, that the different types of information (that is, whether symbolic or indexical) encoded by morphological variation in the case system can also affect whether a language is likely to lose, maintain, or propagate inflexional case-marking.

6.

THE DEVELOPMENT OF THE ITALIC AND FINNIC CASE SYSTEMS

6.0 SUBSTANCE ENTITIES, SCHEMATIC PATTERNS, AND THE DEVELOPMENT OF CASE SYSTEMS

This chapter analyses the roles played by substance entities and schematic patterns in the developments of the Italic and Finnic varieties under consideration. My hypothesis is that there is an implicational relationship between the type of inflexional structure most commonly found in the morphological case system (that is, whether it ise characterized mainly by sub-word substance entities or mainly by schematic patterns), and the developments that occur through time. Proper understanding of different types of morphological variation and the ways they are exploited in the production of inflected forms is therefore essential to explaining the diachronic developments exhibited by the Italic and Finnic case systems.[1]

Since substance entities are largely invariant, speakers may reproduce these forms whole, at both the level of the word and below, and may therefore rely on storage in addition to the analogical extension of schematic patterns (the latter of which can prove to be weaker diachronically than the storage of forms). The proposition put forward here is that the presence of sub-word substance entities therefore provides a language-user with an additional means of producing novel inflected forms, offering a case system (or, at least, certain case-forms) significantly greater diachronic stability than case systems which lack sub-word substance entities. These hypotheses will be explored in reference to the languages introduced in chapters two and three, and explanations for the loss, maintenance, and development of new inflexional case suffixes will be offered.

[1] By extension, there is, I contend, an implicational relationship between the synchronic size of the morphological case system and the types of inflexional variation a language is characterized by. This will be discussed further in chapter seven.

6.1 MORPHOLOGICAL PATTERNS

Identifying different morphological patterns and recognizing how language-users produce novel forms is essential to understanding the development of the Italic and Finnic case systems. Figure 6.1 is intended to help clarify the following discussion:

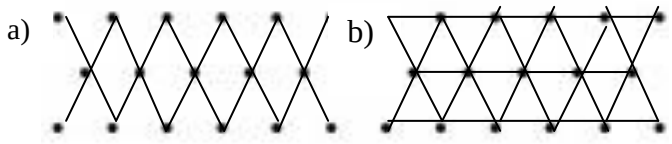


Figure 6.1: Visual representations of morphological patterns

In figure 6.1, the dots are spaced in such a way as to allow multiple ways of linking them by means of straight lines which form uniform patterns - only two (a and b) are given here. Importantly, the same pattern may be drawn in different ways. In example a), one means of achieving the pattern is to draw long, straight, diagonal lines, but an equally valid way to arrive at the same surface pattern is to draw diamonds shapes, or even zig-zags, the peaks and troughs of which touch. In a similar way, language-users may utilize different implicational patterns which nevertheless produce the same results.[2] Straight lines in figure 6.1 are used (as opposed to curved lines, for instance) to represent the most direct means of linking dots. In human pattern-identification, since we are lazy (or perhaps efficient) creatures,[3] we will usually seek and employ the simplest patterns which nevertheless provide us with sufficient information in solving the Paradigm Cell-Filling Problem.[4] The more direct the relationship between the dots, moreover, the clearer the informative patterns, which may aid in their extension across the language.

[2] This will be important for later discussion, particularly on the 'priming' of inflexional patterns in Latin (§6.2), and in reference to the development of novel inflexional case-forms in Veps (§6.4).
[3] Language-users obey the economy principle (or principle of least effort). Zipf (1949).
[4] This is equivalent to the 'fast' thinking of Kahneman (2011). The equivalent of 'slow' thinking may be the analysis of similar patterns by linguists who may posit 'underlying' forms and complex, *ad hoc* phonological explanations for the 'derivation' of one inflected form from another.

For the sake of the following discussion, I will liken the dots in figure 6.1 to substance entities (both at the level of the word and below) and the lines to schematic patterns (that is, the implicational relations that exist between the dots or substance entities). Although the substance entities are (relatively) fixed, the ways in which they are linked or associated in informative ways can differ even within a language depending on various factors that will be discussed in more detail throughout this chapter.

Figure 6.1 is intended to represent visually the way in which a case system 'works'. For instance, in Latin, fully-inflected wordforms (the dots) are inter-connected by means of implicational relations (the lines) in patterns which language-users may identify on consideration of other forms they have encountered. These patterns provide a means of producing novel inflected forms in an informed way, thus reducing the entropy of an unknown form in a lexeme's paradigm.[5] For instance, knowledge of two or more inflected forms of the same lexeme will provide a language-user with a number of dots equal to the number of wordforms known. The language-user then 'draws' a line linking them. With further exposure, the language-user may start to fill more (although, importantly, not necessarily all) cells in an inflexional paradigm, thus building up the number of dots between which linking patterns may become more informative as they begin to provide a language-user with surer means of deducing unknown forms of nominals.[6]

For the purposes of later discussion, I identify three different 'areas' in an inflexional case system: the 'locus of informativeness', the 'locus of uncertainty', and the 'locus of predictiveness'. These loci are found in all languages, although the units which may be described as the 'loci of informativeness' and the 'loci of predictiveness' often change cross-

[5] N.B., the dots are not informative in isolation: it is only their relationship to other dots which permits a pattern to be identified and for informative implications to be utilized.
[6] This description is intended as an approximate metaphor for the processes involved in the production of novel forms, and is not to be taken literally.

linguistically; the locus of uncertainty correlates in all languages with the area of formal variation (which can be the shape of the stem, the inflexional exponent, or even the whole wordform, for instance). I will first introduce and define these different loci and then outline how they interact in language change:

- Locus of informativeness: the level at which the most stable relation of form and function may be identified. This is usually the fully-inflected wordform, although smaller units, such as suffixes or stems, might also reduce the entropy of producing a novel form.
- Locus of predictiveness: this is the level that may be used in the production of a novel inflected form. In languages such as Romanian, it is often the paradigmatic distribution of fully-inflected wordforms; in Latin, the abstract relations that exist between two or more inflected forms; in Finnish, the direct mapping of form and function exhibited by sub-word substance entities. As form-storage is considered a locus of predictiveness, a language will always be characterized by two or more of loci of predictiveness.
- Locus of uncertainty: this locus coincides with the part of an inflected wordform which exhibits inflexional variation (e.g., the stem, the suffix, or the whole wordform). This introduces uncertainty in the production of a novel inflected form. There are thus two different types of loci of uncertainty:
    - the cross-paradigmatic locus of uncertainty, which is where formal variation occurs across different declension classes;
    - the intra-paradigmatic locus of uncertainty, which is where formal variation occurs within the paradigm of a single lexeme (or nominal type).

  Only the first of these will be discussed here and will be referred to simply as the 'locus of uncertainty'.

Let us take Latin and Finnish as examples to illustrate why identifying and understanding these different loci are important. In Latin, the locus of informativeness is the fully-inflected wordform; there are two loci of predictiveness: the fully-inflected wordform (where it has been met previously and stored) and, in the production of novel inflected forms, the abstract relations that exist between known fully-inflected wordforms in the paradigm. The locus of uncertainty is most frequently the suffix (although in certain third declension nominals we find some instances of morphological variation, and hence of uncertainty, in the shape of the lexical stem). Errors are likely to manifest in the suffix, but incorrect implicational patterns are 'selected' at the (non-minimally-entropic) locus of predictiveness, i.e., the implicational relations that exist between fully-inflected wordforms.

In Finnish, by contrast, there are two loci of informativeness: the fully-inflected wordform and, in the majority of instances, the sub-word case and number suffixes. There are also multiple loci of predictiveness: the fully-inflected wordform (as in Latin, where forms are stored); abstract patterns which exist between fully-inflected wordforms (which are particularly informative about the shape of the lexical stem); and, unlike in Latin, sub-word case and number suffixes. There is thus an additional level of predictiveness in the case system of Finnish when compared with that of Latin. The locus of uncertainty in Finnish is most frequently the lexical stem - or, as was discussed in §5.2.2, the 'link' between the lexical stem and the inflexional affix - and this is often where errors manifest. In a similar way to Latin, therefore, we find mistakes occurring in the locus of uncertainty as a result of misapplied implicational patterns.

Certain aspects of the Latin and Finnish case systems are represented visually in the following figures:

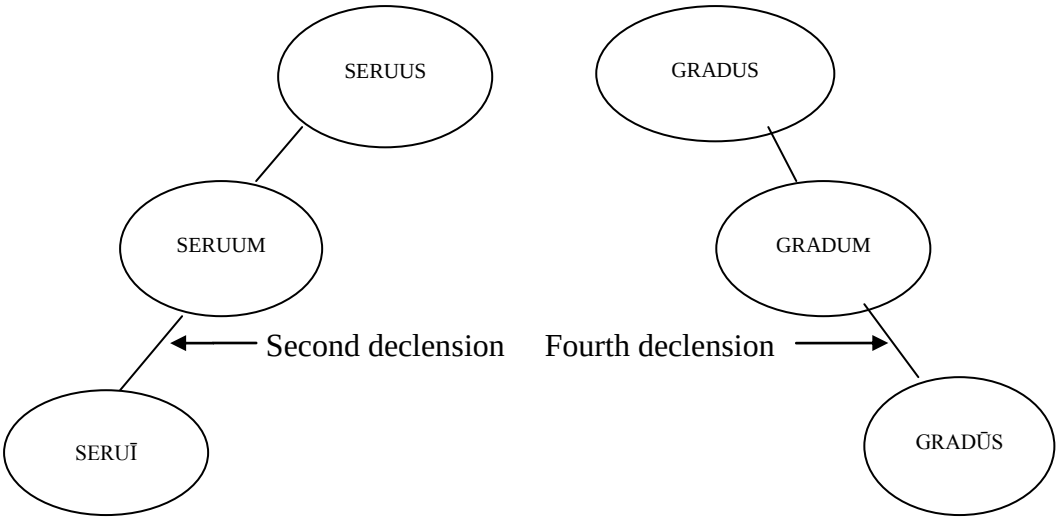


Figure 6.2: Implicational relations in the Latin case system

In each instance in figure 6.2, the nominative singular form is given on top, followed by the accusative singular form on the next level down, and finally the genitive singular form on the lowest level. Here, one pattern holds between one set of three case/number forms (those of {SERUUS}) and a distinct pattern holds between the second set of three case/number forms (those of {GRADUS}). The abstract relations - represented by the lines linking the word bubbles - determine whether a nominal belongs to the second or fourth declension. However, knowledge of only the nominative singular and accusative singular forms of a second or fourth declension noun is insufficient to infer the remaining forms in the inflexional paradigm:

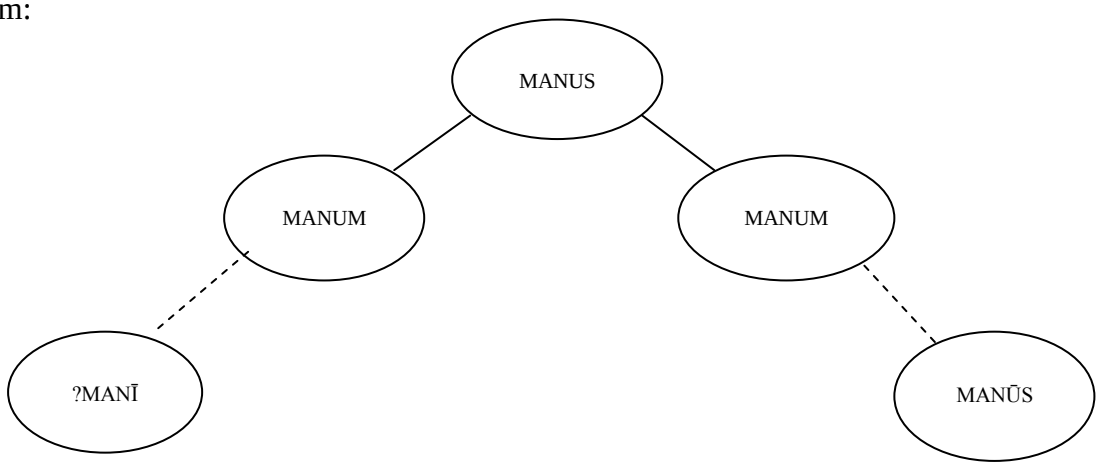


Figure 6.3: Potential inflexional errors in Latin {MANUS} ‘hand’

This contrasts with (a small part of) the Finnish case system:

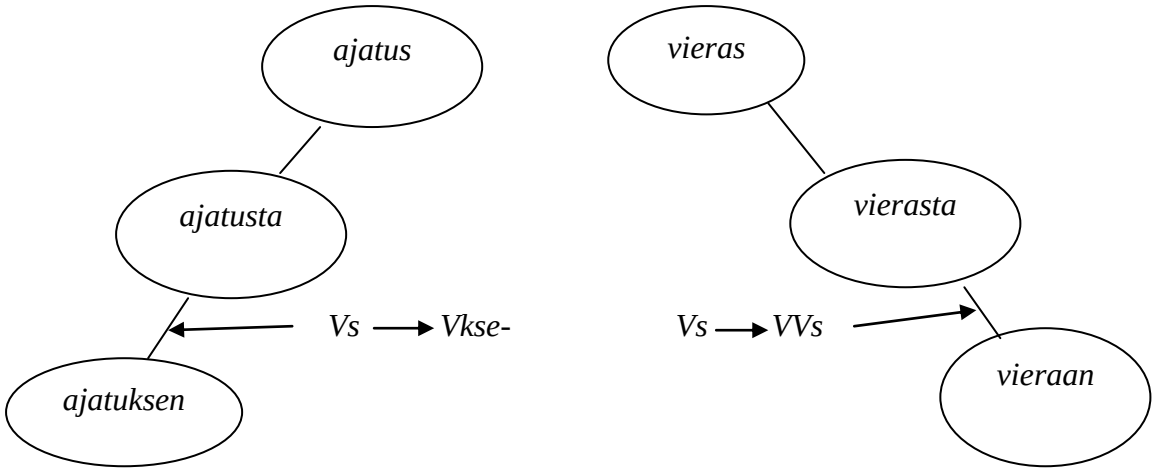


Figure 6.4: Implicational relations in Finnish *s*-stem nominals

In each instance in figure 6.4, the nominative singular form is given on top and is followed by the partitive singular form, and by the genitive singular form on the lowest level. There are three ways in which *s*-stem nominals can inflect (§3.1.1), two of which are included in figure 6.4, and knowledge of the nominative singular and partitive singular forms of an *s*-stem lexeme (two of the most frequent case-forms in the language) is not always sufficient to determine which of these inflexional patterns a nominal will follow.

In the Finnish case system, however, an additional locus of informativeness can be identified:

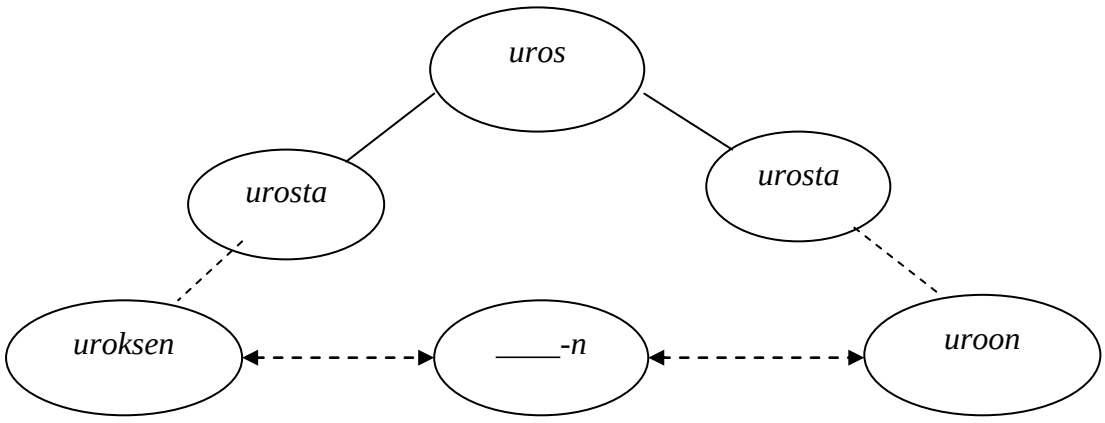


Figure 6.5: Potential inflexional errors in the lexeme {UROS} 'male; man'[7]

[7] Both these inflexional patterns are acceptable for this lexeme.

Given sufficient exposure to inflected forms of a variety of lexemes, a language-user recognizes that genitive singular nominal forms, without exception, are characterized by the structure [STEM + n].[8] In this way, knowledge of other genitive singular forms offers language-users the opportunity to extricate this sub-word substance entity and utilize it in the production of a previously unencountered genitive singular form, even if the expected stem shape is not produced. In this way, where there is not sufficient information to deduce the correct stem shape of a lexeme, it may still be possible for a speaker to produce *a* form, which may be incorrect but is still likely to be interpreted correctly by the hearer.[9]

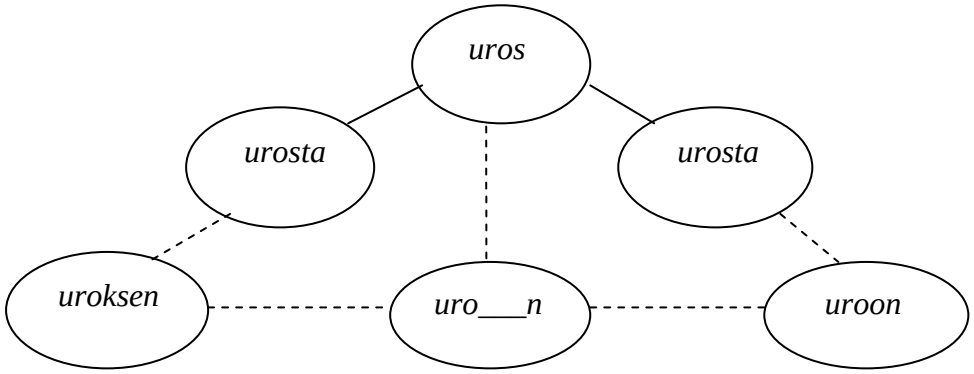


Figure 6.6: Multiple loci of informativeness in the Finnish case system

In Finnish, the sub-word inflexional suffix is often a locus of informativeness. Although these sub-word substance entities cannot always be concatenated straight onto a stem shape and result in the 'correct' surface form, they still provide the language-user with more relevant and reliable information than is found at the sub-word level in Latin. I suggest that where there are more loci of informativeness, a language-user has access to more information in the production of a 'correct' novel inflected form - or, at least, *a* form, the functions of which can be deduced with relative ease by a hearer - than when there are fewer loci of informativeness.

[8] Supported by the L1 and L2 data offered in chapter five.
[9] This reflects the 'probabilistic' nature of language use. Bod (2003), Baayen (2003:233).

The basic claim for which the rest of this chapter aims to provide evidence is as follows: schematic patterns in inflexional systems are very frequently synchronically stable, as demonstrated by the fact that the case systems of the languages discussed here were (or are, for those languages which are still natively learned) at least understood, if not utilized, by language-users. However, those languages which have lost their case systems are all characterized by schematic patterns *alone*, where sub-word substance entities cannot be identified and where concatenation of (possibly incorrect) stem shapes and affixal material cannot reliably produce forms which may be correctly interpreted by a hearer. Schematic patterns are susceptible to disturbance particularly as a result of phonological change.

Where a language is characterized by an abundance of sub-word substance entities, on the other hand, a language-user does not necessarily have to rely solely on the implicational relations that exist in the language, but may also exploit minimally-entropic sub-word material in the production of novel inflected forms. This does not mean, of course, that case systems characterized by an abundance of sub-word substance entities will always maintain or propagate morphological case-forms (for instance, Livonian is likely to have been characterized by more agglutinative structure at an earlier stage, and yet it has lost case distinctions through time). I suggest instead that they are *more likely* to maintain and/or innovate inflexional case-forms through time than those languages in which there are no sub-word substance entities to facilitate the production of novel inflected forms.

These claims will be discussed in more detail in the next three sections. First, §6.2 will explore the developments found in the case-reducing languages Latin, old Gallo-Romance, and Livonian; in §6.3 the maintenance of case in Finnish and Romanian will be outlined and, finally, in §6.4, I will discuss Veps as a case-increasing language.

6.2 CASE-REDUCING LANGUAGES

Of the languages discussed in this study, Latin, Livonian, and the Gallo-Romance varieties old French and old Occitan are described as 'case-reducing' languages. It has already been made clear that the loss of morphological case-forms in these four varieties cannot be explained by appealing to phonetic, analogical, syntactic, or contact-induced developments, nor even a combination of them. Since previous explanations have failed to account satisfactorily for the developments found in these languages, I will here discuss the role of substance entities and schematic patterns and assess how they affect the ways in which case systems develop. In keeping with an evolutionary perspective, it is argued that these morphological structures and the ways they are utilized in the production of previously-unencountered forms must be considered as part of a complex and inter-dependent system of developments, and that language-users are the origins and drivers of the changes found in any given linguistic variety.

First, I will discuss Latin, which is characterized entirely by schematic patterns; there are no sub-word substance entities that can be identified. Fully-inflected wordforms (word-level substance entities) are the most informative elements for language-users. The abstract patterns which exist between these units are what provide a speaker with the means to produce novel inflected forms. Where there is opacification in these implicational relations, it may result in errors in the production of novel inflected forms; these errors usually occur at the locus of uncertainty - that is, the suffix.

I have already suggested that the similarity between the forms encoding nominative singular and accusative singular functions in second and fourth declension nominals encapsulate the sort of uncertainty present in the Latin case system. Where a speaker has met these forms, there is not sufficient information to deduce with certainty the remaining forms

in the lexeme's paradigm. This, therefore, may be analysed as an instance of potential 'instability' in the Latin case system, where it is possible that the implicational relations between two inflected wordforms will not necessarily lead to the deduction of the expected target form. However, it is unlikely to have posed a significant problem: knowledge of only one other form in the paradigm would have permitted a language-user to correctly assign the two lexemes given in figure 6.2, for instance, to the correct inflexion class. This sort of rare occurrence of uncertainty is therefore not sufficient to explain the loss of the Latin case system, although it does present an example of the relative ease with which schematic patterns can be 'interrupted', and hence changed. This is where the discussion on patterns given in §6.1 becomes very important. Recall that the dots (i.e., the substance entities), which are usually those forms which are stored and reproduced whole, are static. Schematic patterns, on the other hand, might differ from individual to individual, since there are many different ways to link the dots in 'patterned' ways. Where there is the possibility for different interpretations of the relationship(s) between static substance entities, there is greater possibility for change. This was exemplified in figure 6.1 not only in the different patterns that were identified, but also in the discussion following on the different ways in which the 'same' patterns could be achieved.

In Latin, similarly, the same morphological patterns exhibited by nominals may be 'drawn' in different ways. For instance, both the following analyses of certain third declension nominals are valid:

re:ks	vs.	re:k + s
[FULLY-INFLECTED WORDFORM]		[LEXICAL STEM] + [SUFFIX]

Figure 6.7: Analysing Latin nominative singular RĒX

Because this same suffixal shape *-s* (whether represented orthographically or not) is found in many other third declension nominals, this may be analysed as a suffix which can be associated with nominative singular functions, at least for certain lexemes belonging to the same inflexion class. Equally, however, because many other third declension nominals differentiate the 'nominative singular' functions by means of a distinct stem shape which does not (necessarily) end in a sibilant, other language-users may analyse these forms as encoding nominative singular functions by virtue of the different stem shape alone, and not recognize any sort of suffixal material. The result is the same: the form RĒX 'king' is analysed as the nominative singular form, but the analysis of the ways in which this form 'relates' to other inflected forms in the language can differ.

It is possible that 'priming' effects alter the patterns that language-users identify and exploit in the production of novel forms. For instance, it may be that, when a Latin language-user encounters the nominative singular form of a third declension nominal not characterized by a final sibilant (e.g., NŌMEN 'name'), they may be more likely to analyse another third declension nominative singular form as encoding these functions by means of a different stem shape. On the other hand, when they have recently encountered other third declension nominative singular forms ending in *-s*, they may have been more likely to analyse the final sibilant as the 'marker' of nominative singular functions. To continue the metaphor given in §6.1, in this analysis, the dots remain in place, but the ways in which informative lines are drawn differ depending on exposure to previously-encountered patterns.

Regular sound changes can also affect, sometimes drastically, the means that language-users employ in the production of novel forms. For example, certain phonological developments described in §2.1.2 led to the syncretism of the nominative, accusative, and ablative functions in the singular of first declension nominals in Latin. As a result of these phonological developments, the predictability of form actually becomes greater: in these

lexemes, knowledge of the nominative singular form would have provided sufficient information to predict the forms of the accusative singular *and* the ablative singular of the nominal. In this way, phonetic erosion would have in some ways *strengthened* the implicational relations between the formal expression of particular functions, since invariance is one means of achieving perfect predictability (where the lexical item is known).

Following the emergent syncretism of the nominative, accusative, and ablative in the singular of first declension nominals, there are no instances of the re-establishment of morphological distinctions. In some languages, such as Estonian, sub-word elements spread throughout the language in order to re-establish transparent sub-word structure where it has been lost as a result of phonetic erosion. For instance, although many nominals are characterized by a 'short' illative singular form which cannot be decomposed into 'lexical stem' and 'inflexional material', the generalizaiton of the element *-sse* reverses the syntagmatic opacity resulting from regular phonological developments. Nominals can now also have a 'long' illative singular form with the structure [[GEN.STEM] + [*sse*]]:

Singular	**{MAJA} 'house'**	**{RIIK} 'state, country'**	**{INIMENE} 'person'**
Nominative	*maja*	*riik*	*inimene*
Genitive	*maja*	*riigi*	*inimese*
Partitive	*maja*	*riiki*	*inimest*
Illative$_1$	*majja*	*riiki*	*inimesse*
Illative$_2$	*maja-sse*	*riigi-sse*	*inimese-sse*

Table 6.1: Estonian 'second' illative forms

In Latin, however, we find no such development 'reversing' the syncretism of the nominative, accusative, and ablative in the singular of first declension nominals. For example, there are no instances of the second (or fourth) declension accusative singular ending *-u* (following the loss of final *-m*) being extended to differentiate nominative singular from accusative singular (i.e., accusative singular **?*puellu* 'girl'). Here, I suggest, is where one of the diachronic 'weaknesses' of a case system characterized by a lack of sub-word

substance entities is most obvious. Implicational schematic patterns in Latin exist not only between two or more forms, but are also part of specific inflexion classes. Suffixal alternations which encode indexical information (§5.2) (and so provide information not only case/number distinctions, but also about the inflexional behaviour of a lexeme, and sometimes its gender, for instance), are much less likely to be 'transferred' from one declension to another, since they are only informative in relation to the other suffixal shapes that are found in the paradigm of a lexeme. That is, a form ending in *-u(m)* can only be interpreted as accusative singular because of its relation to a form ending in *-us* which encodes nominative singular functions (and, of course, its relation to other forms in the paradigm). This form *-u(m)* cannot therefore be introduced into the inflexional paradigm of a first declension lexeme, since the nominative singular form is characterized by a form ending in *-a*. Even in adjectives, where forms such as BELLA and BELLUM are found, there is no indication that suffixes belonging to distinct inflexion classes are 'borrowed across' to mitigate the effects of phonetic erosion. The suffixal alternations, often diagnostic of declension, are involved in cohesive, indexical networks of interpredictability, but they are not informative outside of the inflexion class to which they belong.

There are a couple of exceptions to these observations. In the first place, there appear to be sporadic attestations of third declension lexemes exhibiting a dative/ablative plural form in *-īs* (or sometimes *-es*) where *-ibus* is expected.[10] In the second place, third declension masculine nominative plural forms are sometimes reconstructed which end in *-ī*, the nominative plural exponent found in second declension lexemes (rather than *-ēs*), particularly on the evidence available from old French and old Occitan.[11] In these instances, we appear to find single

[10] Third declension lexemes with a dative/ablative plural form in *-īs* are also sometimes suggested to be reflected in Romanian (and Italian) lexemes etymologically belonging to the Latin second declension, discussed briefly in §2.3.2 in reference to the development of the Latin lexeme {LĒX} 'law' (table 2.17).
[11] Pope (1952:309); also discussed in §2.2.3.

suffixes which have been 'transplanted' from second declension to third declension lexemes, contrary to what was claimed above.

In light of evidence such as that presented by Gaeng (1984), it is not clear that the occasional 'replacement' of dative/ablative plural forms in *-ibus* with ones ending in *-īs* was actually the result of the analogical extension of the second declension exponent to third declension lexemes. First, attested dative/ablative plural forms are relatively rare, having been replaced generally with prepositions in addition to a nominal form inflected in the accusative. Second, instances in which forms ending in *-īs* are not always unambiguously ablative. For example, a Christian inscription reading *ex equitibus siculares* (=*singulares*) from Dalmatia may exhibit an analogically-extended second declension ablative plural exponent, or it might be an example of an accusative plural with the preposition *ex*.[12] The inscriptional evidence is by no means unambiguous, but it does appear as though there was fluctuation in the use and forms of the dative and ablative in the plural of Latin lexemes generally. Gaeng surmises that: "[t]he examples in which an ablative in *-ibus* occurs side by side with forms in *-es* or *-is* in the same function seem to reflect its continued use on the formal level, while pointing in the direction of a replacement by a *casus generalis* on the level of speech."[13]

If the presence of dative/ablative plural *-īs* in certain third declension nominals is assumed to be an analogical development, the theory posited above may be altered. It may instead be suggested that, where a form is not diagnostic of the inflexion class to which a nominal belongs, it may be more likely to get extended to discriminate the same case/number values in different declension classes. The dative/ablative plural form *-īs* is characterized by a lighter indexical load than, for instance, a nominative singular form in *-a* (which signals

[12] Gaeng (1984:84f.)
[13] 1984:85.

unambiguously that the lexeme follows the inflexional pattern of first declension nominals), and does not provide sufficient information to determine how a lexeme inflects (that is, whether it inflects like first or second declension lexemes). Notably, the form *-ibus* is similarly uninformative about whether a lexeme behaves as a third or fourth declension nominal, but since fourth declension lexemes were almost entirely attracted to the second declension, it is possible that the inflexional ambiguity of the dative/ablative plural exponent *-ibus* was gradually lost along with the separate fourth declension inflexion class.

Such an explanation does not work for the nominative plural exponent *-ī*, however, which is a diagnostic form - that is, it provides nearly minimally-entropic indexical information about the remaining forms of a lexeme.[14] It could be that this form was analogically extended in order to encode a morphological difference between nominative and accusative case values in the plural of third declension nominals, but such an explanation works counter to the extension of the accusative plural marker *-ās* to encode nominative plural functions (§2.2.3): how could the third declension pattern of syncretism be analogically extended to first declension nominals whilst there is a simultaneous morphological distinction of these two sets of functions in development?[15]

Upon closer consideration of the available data, however, it appears as though the supposition that there was a spread of nominative plural *-ī* to etymologically third declension nominals might be based on theory rather than attested evidence. This development is hypothesized to have occurred to account for the lack of etymological *-s* in the nominative plural forms in old French and old Occitan nominals which were historically from the Latin third declension. However, as was discussed in §2.2.3, the lack of etymological *-s* in these

[14] 'Nearly' minimally-entropic because there are a handful of Greek loanwords in the second declension which exhibit slight inflexional differences to native second declension lexemes (as discussed in §2.1.1).
[15] This is reminiscent of the drawbacks in reference to appealing to analogical explanations of certain developments offered in §2.1.3.

nominative plural forms was not necessarily the result of the analogical extensions of final *-ī* from etymological second declension lexemes. Instead, the lack of suffixal material in certain forms might be the result of the pattern of morphological variation in the plural exhibited by *murs*-type nominals.[16] Do we therefore need to posit the analogical development of nominative plural third declension lexemes *-ēs* > *-ī* at all?

The evidence from Romance on this thorny issue is inconclusive, since modern nouns and adjectives largely continue the accusative form from Latin. In Italo-Romance and Romanian, third declensions plural forms in *-i* have developed not from the analogically-extended second declension exponent *-ī*, but from etymological *-ēs* via **-e̯i*.[17] The inscriptional evidence in support of an analogical extension of second declension nominative plural *-ī* also appears to be lacking. In Gaeng's (1984) data set, for instance, no third declension lexeme exhibits a nominative plural form ending in *-ī*; forms end either in *-es* or, more sporadically, in *-is*. The main evidence pointing towards the analogical extension of second declension nominative plural *-ī*, therefore, is from reconstructive evidence of the old Gallo-Romance case system. If we posit such a development, therefore, it is on the strength of theory, and not of observable data.

If there was an analogical spread of second declension *-ī* to third declension nominals with no concurrent shifting of inflexion class, then this example runs counter to the hypothesis offered above, that the nature of the Latin schematic patterns prevents the generalization of (particularly indexically class-diagnostic) exponents across the language. It may indicate that, through time, the cohesiveness of the major inflexion classes in Latin weakened so that even

[16] Such is the relative timing implied by Maiden (2011a:165), who notes that "the zero inflection in the nominative plural [of *murs*-type nominals] reflects historically regular deletion of unstressed final -i. This pattern was frequently extended to original third declension masculines".

[17] Maiden (2011a:164). Even if the Dalmatian plural form in *-i* has not developed from *-ēs* but from the analogical extension of the nominative plural *-i* found in demonstrative and pronominal forms (historically descended from Latin ISTĪ and ILLĪ) after Maiden (to appear), this generalization is one that has happened in Dalmatian and not one that occurred in Latin.

diagnostic suffixes could be transferred across declensions. Further research in this area is required; reference to the development of the case systems of other languages, particularly those characterized by tight networks of indexicality and interpredictability, might provide similar examples of the Latin second declension nominative plural suffix *-ī*, and offer insight into its generalization.

Despite these few examples, in many ways, schematic patterns in Latin appear to be relatively 'rigid', in that they frequently cannot be transferred across declensions (although paradigmatic schemata may be analogically extended from one declension to another, as potentially seen in the development of first declension nominative plural forms in *-ās*). Where case distinctions are opacified by regular phonological developments, therefore, it is not possible for sub-word substance entities, since there are none in Latin, to 'replace' the lost suffixal distinctions. Schematic patterns are usually bound by inflexion class and, although there is evidence of a great deal of shifting of lexical items to new declensions, we do not as frequently find suffixal alternations re-establishing a distinction between certain case/number values which may no longer be discriminated morphologically.

Syncretism is extensive in the Latin case system, and language-users were likely to be familiar with many other instances in which multiple functions were encoded by a single form, even where the same functions are formally distinct in other declensions. The development of new instances of syncretism therefore does not introduce any sort of morphological aspect that is 'unexpected' or unfamiliar to the case system of this language.[18] Although, therefore, the instance of syncretism in the singular of first declension nominals may appear problematic on paper (where the formally distinct expression of very different functions has been effaced), it would very rarely have led to uncertainty in the syntactic

[18] Syncretism may therefore be referred to as an instance of 'system-dependent naturalness' from a Natural Morphology perspective. Wurzel (1987).

context, and is perfectly common in the inflexional system. This may present another reason why no 'new' morphological means of differentiating the nominative, accusative, and ablative functions are introduced in first declension singular nominal, since syncretism was found elsewhere in the system.

Although the schematic patterns that exist between fully-inflected wordforms are the means that language-users exploit in the production of novel inflected forms in Latin, these abstract relations are not always minimally-entropic in the same way as sub-word substance entities. This means that there is still some uncertainty inherent in the production of novel forms in languages characterized by schematic patterns alone, and is thus often why we find 'errors' in inflexion. Where phonological changes opacify older implicational relations through time, this does not mean the *loss* of schematic patterns, but the *shifting* of informative abstract relations. In the first declension, uncertainty associated with formal variation exhibited by the suffix was eliminated by syncretism. More certainty is associated with invariance than with morphological difference, and has become almost universally established in Romance.

It was not simply because Latin was characterized solely by schematic patterns that it lost its morphological case system in nominals. Indeed, as will be discussed in chapter seven in reference to the Latin verb system, many of the schematic patterns of the Italic verb have proven to be very stable through time. Instead, a convergence of factors aided in the shifting of the locus of predictiveness - that is, from implicational patterns between fully-inflected wordforms to stored forms that were invariant for case. Latin was characterized by indexically-informative formal variation of the suffix, particularly between the most frequent case-forms of lexemes, which provided the language-user with often sufficient information to deduce most if not all other forms in the inflexional paradigm. Through time, phonological developments led to the syncretism of certain forms and, because the schematic patterns in

Latin held most cohesively within the inflexion class of a lexeme, these emergent syncretisms were not 'undone' - that is, morphological distinctions were not reintroduced in order to formally differentiate particular functions. Greater predictiveness was afforded by invariance at the word-level.

Moreover, alongside the shifting locus of predictability towards invariance for case was an attendant preference for the analytic expression of particular functions once encoded by bare case-forms. These prepositional means of discriminating case functions were also minimally-entropic, and so invariance at the level of the word generalized. The rise of configurationality, found not just in the noun phrase but throughout the Italic languages, was an intertwined development, neither caused by nor causing the loss of the case system, but moving alongside it, both influencing and being influenced.

The evidence from Latin suggests that even a stable case system that is apparently quite cohesive synchronically, if not also characterized by sub-word substance entities, is more likely to become unstable over time (as the result of phonological, analogical, functional, syntactic, and contact-induced changes, for instance) and to lose (at least some, if not all) case distinctions. The uncertainty associated with schematic patterns, particularly where they are 'interrupted' by phonetic erosion or analogy, for instance, and the lack of means to re-introduce morphological distinctions because of the indexical nature of formal variation in the suffix, are likely to lead to the gradual demise of the system.

In a similar way to Latin, phonological developments in old French would have opacified formal distinctions between different case/number functions. In the majority of lexemes, the loss of the final sibilant before consonants would have led to perceived lexical invariance, since the sibilant *-s* is found across the language, regardless of declension (although the affricate *-z* is also found). The same observations do not hold, however, for old Occitan, as

was discussed in §2.2.2. How, then, do we account for the loss of case/number distinctions in this variety?

It is clear that, the more phonologically similar the means of differentiating case(/number) functions across different declensions, the less indexical information is encoded by suffixal material; in old French and old Occitan, suffixal alternations are less highly discriminated, and provide much less indexical information, than the suffixal alternations in Latin. In this way, a fully-inflected wordform in old Gallo-Romance, met in context, offers less information about the paradigmatic inflexional patterns it follows. Thus, a novel nominal, in addition to the context in which it occurs, is not sufficient to permit the speaker to produce any other inflected forms in the lexeme's paradigm with perfect certainty. In Latin, on the other hand, in many instances a single form of a novel lexeme would have provided (near) minimally-entropic information about the remaining forms in its paradigm.

In old Occitan (and, indeed, in old French), only knowledge of at least one other inflected form of a lexeme (and, more frequently, two) provided the language-user with sufficient information to deduce the declension to which the nominal belonged, and thus how it behaved inflexionally. In this way, the nature of the schematic patterns found in old Gallo-Romance indicate that this system was likely to be a lot less stable diachronically (and, as indicated by the fluctuations in the distribution of distinct forms discussed in §2.2.3, perhaps also synchronically) than the Latin case system. Even though, therefore, the old Occitan case system did not lose its distinctions as the result of phonetic erosion, the amount of information provided by morphological alternations in nominals was insufficient to deduce their inflexional behaviour exhibited, which aided the demise of the case system by not

providing consistently (near-)minimally-entropic information for the production of previously-unencountered forms.[19]

Finally, Livonian is here considered to be a case-reducing language, although it contrasts with Latin, old French, and old Occitan in that it has not entirely lost its case system. Moreover, Livonian has developed a novel dative form even alongside the loss of the external local case-forms, despite being characterized by a great deal more fusional structure than either Finnish or Veps. In Veps, as will be discussed in §6.4, one means of (re-)establishing morphological case distinctions is by the suffixation and grammaticalization of postpositions (or other postpositional material). This is, in part, because of the behavioural similarities of sub-word and word-level substance entities: a language-user recognizes that they each occur in the same place relative to the nominal (i.e., following it) and that both are invariant across number and lexeme, for instance. Thus, the presence of sub-word substance entities facilitates the reanalysis of postpositions into suffixal markers.

Livonian, however, does not appear to have developed this way as,[20] according to the work conducted on the development of the dative (outlined in §3.2.2), the modern form in *-n/ -õn* descends from pre-existing suffixal material (whether from the genitive, the essive, or a lative marker). Livonian is not characterized by an abundance of sub-word structures, as was demonstrated in chapter four; instead, suffixes cumulatively express case and number, exhibit morphologically-distributed, phonologically-distinct exponents, and are characterized by a degree of fusion with the lexical stem that sometimes makes it difficult to separate the two. Thus, case suffixes and postpositions in this Finnic variety do not exhibit the same sorts of behavioural similarities as are found in Veps. The development of the Livonian dative is instead an example of 'exaptation' (after Lass 1990 and 1997) of material once employed to

[19] In addition to phonetic erosion, this lack of informativeness is likely to have contributed to the loss of the old French case system.
[20] One exception may be the development of the translative-comitative form, discussed in §3.2.2.

discriminate genitive, essive, or lative functions.[21] The phonological material remains, but the ways in which it is used changes (or 'adapts') through time. In this way, the behaviour of the dative more nearly reflects the behaviour of other inflexional suffixes than the postpositions in the language: it cumulatively expresses number and case, and exhibits formal variation across declension (albeit phonologically- rather than morphologically-conditioned) across number and lexeme (although it is much more easily separated from the lexical stem than some other suffixal markers).

The dative provides only a single example of an innovative case-form in Livonian. Otherwise, phonetic erosion has largely had the effect of certain case distinctions or certain case suffixes being lost. Two main developments are discussed here: the loss of what originally might have been described as a sub-word substance entity (the genitive singular marker *-n*), and the loss of the external local cases. The first of these developments has not led to the loss of the genitive value, despite the value no longer being discriminated by means of a suffix, and even though the genitive is only phonologically distinct from the nominative in a few lexemes, and only in the singular. The predictive information once, theoretically, provided by a sub-word substance entity has been lost, and now schematic patterns must be used in the production of novel genitive singular forms. This provides further evidence for the earlier claim that there is no system which is characterized solely by sub-word substance entities, but that schematic patterns are always found in a case system. With the loss of sub-word substance entities, there already exist alternative means of deducing previously unencountered inflected forms. The relationship that the whole genitive singular wordform bears to other fully-inflected forms in a lexeme's paradigm provides more information in the

[21] Exaptation is here defined as "the co-optation during evolution of structures originally developed for other purposes" after Lass (1997:316). There is one slight difference between my use of the term and the way in which it is used elsewhere. Lass (1990 and 1997) refers to the exaptation of functionless 'junk' in language, or historical residue that has no function. In contrast, I believe that there is no such thing as 'junk' which is truly functionless. It may be aligned with considerations of style or register, rather than with the distinguishing of case or number values, for instance, but it cannot be described as being without function. The definition of exaptation quoted above can still be used without adaptation even given this slight difference in perspective.

production of novel forms than sub-word substance entities. Finally, the genitive singular is one of the most frequently-used case-forms,[22] which may have aided in its maintenance despite the loss of the sub-word substance entity *-*n*.

In contrast to the development of the genitive singular, although nominals now no longer distinguish the external local case values inflexionally (these functions are now distinguished by means of adpositions), the exponents have not been wholly lost through phonetic erosion and are still found in certain lexemes. The extant texts do not allow us to accurately trace the loss of these values, but I hypothesize that these forms were likely to be less frequently used than the grammatical case-forms (considering corpora counts from other Finnic languages). Without minimally-entropic means of deducing novel forms (for instance, without fully-form or sub-word substance entity storage), the schematic patterns which held between these less frequent forms were not as strong as those which permitted the maintenance of a genitive value. Greater certainty was instead provided by minimally-entropic adpositions. As a result of the paucity of the data, however, such suggestions must remain speculative at best.

Intriguingly, Livonian does not only exhibit the most 'fusional' structure of any Finnic variety, but also maintains the fewest case distinctions. Unfortunately, it is impossible to trace through time whether the number of case distinctions was reduced following (or perhaps, alongside) the phonetic erosion which has led to the more fusional nature of the case system. But it is clear that schematic patterns are often the only means of determining the correct shape of a lexeme in Livonian.

It appears as though schematic patterns, to a large extent, preclude the extension of indexical suffixal markers across lexemes. Instead, we more frequently find the transfer of

[22]It may be that, at least for the most frequent lexemes in Livonian, the partitive singular form is rote-learned and reproduced whole. In the Murdekorpus, 6,038 genitive singular forms are counted. cf. 10,606 nominative singular forms and 2,111 partitive singular forms. http://www.murre.ut.ee/mkweb/. Accessed: 23/10/2018, 14:23.

lexemes from one declension to another; those schematic patterns which permit inflexion classes to be identified by language-users appear to be powerful attractors of lexemes. Sub-word substance entities, on the other hand (as was discussed briefly in reference to Estonian above), tend to *be* attracted. Furthermore, despite instances in which schematic patterns appear to be synchronically cohesive (as in Latin, for instance), this type of morphological variation is more likely to be disrupted by regular developments in a language. An inflexional case system which consists solely or largely of schematic patterns is therefore more likely to lose some or all of its morphological case distinctions through time.

6.3 CASE-STEADY LANGUAGES

In §2.4 and §3.4 respectively, I analysed Romanian and Finnish as 'case-steady' languages. Typologically, the morphological structures of these varieties are very different: Romanian is characterized by schematic patterns, whilst Finnish is characterized by an abundance of sub-word substance entities (in addition to schematic patterns). There is a further important difference between the two languages in the number of morphological case distinctions they maintain: Romanian exhibits the smallest number of morphological case distinctions possible (i.e., just one), which is limited to a subset of nominals. Finnish, on the other hand, maintains twelve productive morphological case-forms, four fossilized values (prolative, comitative, instructive, abessive), and two novel suffixes in restricted usage (excessive, second comitative). The presence and absence of sub-word substance entities is essential to understanding how two case systems which are so different - in terms of their size and the morphological structures by which they are characterized - both appear to have been 'steady' through time.

In Romanian, since there is no certainty afforded below the level of the word,[23] the fully-inflected wordform is here analysed as the main locus of informativeness; these wordforms must be stored and reproduced whole in order to be generated 'correctly'. Where an inflected form of a lexeme is not known, however, the implicational relations that exist between previously-encountered forms can be used to produce novel forms (though these abstract patterns are not minimally-entropic). Finally, the paradigmatic schemata of inflected forms, correlated (to a large extent) with gender, provide reliable information about the distribution of fully-inflected forms where they are known to the speaker. In this way, there are still multiple loci of informativeness in Romanian, despite the lack of sub-word substance entities, but only one of them is minimally-entropic (that of the storage of a fully-inflected wordform); the second (abstract patterns) does not provide minimally-entropic information in the deduction of novel items; the third (paradigmatic schemata) is more informative about the organization of known wordforms in the paradigm than the shape of the inflected lexical item.

If the case system of Romanian is considered a language-internal innovation (as was argued in §2.3.2-3), then we find a development which is similar in some ways to that of Livonian. Namely, that the distinction between nominative-accusative and genitive-dative in feminine singular nominals is the result of the exaptation of pre-existing (morpho-)phonological material, rather than the grammaticalization of a postposition, for example. In this way, the innovative case distinction matches the other discriminative morphological structures exhibited by Romanian nominals (for instance, the fusional nature of lexical stem and inflexional material; formal variation in the stem and suffix; cumulative marking of

[23] There may arguably be certain forms which could be analysed as sub-word substance entities - the plural suffix *-uri*? - but these are few and far between. The fully-inflected wordform is never predictable from the agglutination of sub-word parts, since there are frequently unpredictable alternations in the lexical stem, which occur alongside suffixation of case/number material. Even in the example of the plural suffix *-uri*, there is evidence in Romanian to suggest that this form is not always analysed as a 'plural' marker, but that the final *-i* alone tends to be associated with plurality whilst the formative *-ur-* may be analysed as part of the stem, as part of the plural marker *-uri*, or as a "secondary desinential marker of number." Maiden (2016f:11).

feature values, i.e., number and gender). It might be more accurate to hypothesize, therefore, that languages that lack sub-word substance entities are less likely to develop new case-forms, but that, where they do, the novel desinences must match the morphological structures already found in the language and are therefore more likely to develop from the exaptation of pre-existing formal variation. This would also account for the development of the Courland Livonian dative.

There are a number of similarities between Romanian and the old Gallo-Romance. In all of these languages, only two case values are distinguished morphologically in the nominal system; the case systems of all these varieties are characterized by schematic patterns rather than sub-word substance entities; and the most informative units are fully-inflected wordforms. In light of these similarities, why do we find the persistence of a case system in Romanian but not in old French or old Occitan?

The answer to this has to do, I suggest, with the extra locus of informativeness in Romanian compared to old Gallo-Romance. The additional predictive information provided by the paradigmatic schemata of feminine nominals in Romanian (although it is not minimally-entropic) may aid in supporting a case system characterized by schematic patterns by providing the predictive information not afforded by sub-word substance entities. In old French and old Occitan, where the distribution of fully-inflected wordforms is never predictable, not even from indexical information provided by formal variation in the suffix, for instance, there are fewer entropy-reducing measures available to the language-user in the production of novel forms.

In addition, the morphological variation found in the case system of a Romanian nominal is more discriminative than that found in the old Gallo-Romance case paradigm, and can, in some instances, provide indexical information about the gender of a lexeme. In old French

and old Occitan, for instance, the most common means of differentiating case/number values is that of the alternation of *-s/-z* and *-Ø*. These suffixal alternations do not provide information about, for instance, the gender of a nominal, or about the inflexion class to which it belongs. All the suffix *-s/-z* serves to differentiate is one form from another *within* the paradigm (and, indeed, within the syntagm - Schøsler 1984). In Romanian, on the other hand, certain morphological structures can provide information about the gender of a lexeme. For instance, a form ending in a consonant, a glide, or *-u* in the singular takes masculine agreement; a form ending in *-e* or *-uri* in the plural takes feminine agreement,[24] and knowledge of two forms in the paradigm of a lexeme almost always provides information about the gender of a noun.[25] Since gender is often also correlated with case-marking, gender may also provide some information about the paradigmatic distribution of forms. The shape of the suffix then provides the language-user with some information about whether a lexeme is likely to inflect for case. In this way, where a Romanian language-user encounters a feminine nominal with a nominative-accusative singular ending in *-ă*, they know that the form will inflect for case.[26] The paradigmatic distribution of forms in Romanian - and, therefore, the distinguishing of case in the singular - is not likely to spread, however, for the same reasons. Because this distribution is correlated with morphology (nominative-accusative forms ending in *-ă*) and with gender (feminine), it is unlikely that masculine or *genus alternans* forms will start to exhibit the same distribution of forms, since they have a different gender value (at least in the singular) and are almost never characterized by final *-ă* which is an almost unambiguous marker of feminine gender.

In Romanian, gender and morphology reduce the uncertainty about the distribution of different word-level forms in the paradigm. In old Gallo-Romance, on the other hand, we find

[24] Maiden (2016a:114f.).
[25] Although not always: *vulpe ~ vulpi* 'fox' (feminine), but *şarpe ~ şerpi* 'snake' (masculine).
[26] The certainty of other feminine lexemes inflecting for case is reduced by invariant nominals.

that neither of these considerations may be exploited by language-users in determining whether and how a form inflects for case. That is, morphology is not sufficiently indexical (the same suffix *-s/-z* is found regardless of nominal, although its distribution in the paradigm differs slightly) to provide information about the distribution of forms or the inflexional alternations exhibited by the nominal, and gender does not provide minimally-entropic information about how or whether a lexeme inflects for case.

Since the stability of the Romanian case system appears to come from the perfect informativeness of stored wordforms, the stability of the Finnish case system cannot be accounted for in the same way, since, "[w]ith the combinatorial explosion of possible forms in Finnish, the likelihood of extensive storage of such forms in the mental lexicon diminishes considerably".[27] Instead, in Finnish, there are multiple loci of informativeness. Language-users can utilize not only the fully-inflected wordform (where it is stored) as a minimally-entropic form~function mapping, but also sub-word substance entities which can be associated with certain case (and/or number) values. This provides two loci of informativeness when compared to only one in Romanian, for instance. This also means that there are multiple loci of predictiveness: the fully-inflected wordform (where it has been encountered before), the schematic patterns (which aid particularly in the deduction of the stem shape), and the often invariant inflexional suffix. The first and last of these are minimally-entropic.

In Finnish, therefore, there are more loci of predictiveness than in Romanian, which provide language-users with information - some of which is minimally-entropic - about the production of a novel inflected form, which adds to the stability of the case system. The presence of maximally-predictive sub-word elements enables a language-user to produce a greater number of novel inflected forms within a lexeme's paradigm. Although schematic

[27] Bertram et al. (2000:490).

patterns can be very powerful, there appears to be a 'limit' to the predictive information that they provide a language-user, which restricts the number of case-forms that can be maintained through time in a language which lacks sub-word substance entities.

The evidence presented in chapter five suggests that both implicational information and the storage of symbolic (§5.2) sub-word substance entities are used in the production of novel forms in Finnish. Moreover, it has already been noted that forms that are stored tend to be more stable through time. For this reason, I suggest that sub-word substance entities are likely to be more diachronically stable (although still often sensitive to phonological developments) than schematic patterns, since they can be abstracted away from fully-inflected wordforms, associated with particular functions, and reproduced whole. This is at least part of the reason why the large case system of Finnish has been stable for so long: language-users recognize and utilize the symbolic constancy of sub-word units in the production of novel inflected forms. Since there are more potential forms in the paradigm of a given nominal, more predictive information is required in the 'supporting' of the large case system through time.

Romnian and Finnish, two typologically very different languages, have both maintained their case systems. However, the different structures found in these varieties provide different means for language-users to produce novel forms which, I have suggested, affects the number of case distinctions that can be maintained. In the first place, a language may be characterized by very few word-level substance entities that are stored and accompanied by powerful paradigmatic schemata providing information about the distribution of fully-inflected wordforms. Extra-morphological considerations such as gender may also provide information about which forms are likely to inflect for case. In the second place, a language may be characterized by a 'large' case system which is supported by frequent instances of isomorphic sub-word form~function mapping in addition to schematic patterns and the storage of fully-inflected wordforms (where they have been encountered before). The extra layer of

minimally-entropic predictiveness provides the language-user with additional information that can be used in the deduction of a novel form when compared with those varieties which are not characterized by sub-word substance entities.

6.4 CASE-INCREASING LANGUAGES

Of the languages discussed, Veps is the only one which is clearly case-increasing, since this variety has developed (or, more accurately, is in the process of developing) at least eight new case suffixes. In the southern dialect, these younger forms exhibit a great deal more case-like than adpositional behaviour (§3.3.4). Athough in the written, northern, and central varieties, the forms which have developed from postpositions and postpositional adverbials may not yet be considered 'fully-developed' morphological case-forms, they already exhibit more suffix-like behaviour than the cognate postpositions and adverbial particles in many other Finnic languages.

In order to understand why we find an accretion of morphological case-forms in Veps, attention must be drawn to the behavioural similarities exhibited by synthetic and analytic case-markers in this variety. In the first place, it is perhaps obvious, though essential, to note that suffixes and postpositions occur in the same position relative to the nominal - i.e., following it. Moreover, consideration of the locus of uncertainty - which in Veps is, as in Finnish, the 'link' between the stem and the inflexional suffix - is similarly placed between the nominal stem and the inflexional suffix, as between the nominal stem and the postposition. It is here where formal variation is found in both instances:

6.1.a *sug<u>a</u>-n*
comb-GEN.(SING.)

6.1.b *sug-<u>i</u>-den*
comb-PL.-GEN.

6.1.c **sug-<u>i-den</u> kerta*
comb-PL.-GEN. with

The common stem of the lexeme {SUGA} is *suga-*; this alternates with the plural stem shape *sug-*; the formal variation occurs before the inflexional exponent *-i-* in all plural forms but the nominative. Relative to the postposition, it is the inflexional case exponent which alternates (according to number, for instance). In this way, the locus of uncertainty is the same in respect of the case suffix and the postposition.

Through time, language-users become unsure where the nominal 'ends' and where the postposition 'begins'; the grammaticalization of postpositional material occurs where the uncertainty associated with inflexional variation is found, leading to reanalysis. We thus find the shifting of patterns, where the developments match the morphological structures and properties already found in the Veps case system. To take a visual representation: where originally language-users would have 'drawn' two separate lines in the production of a postpositional phrase, these two lines would have become 'merged' into one:

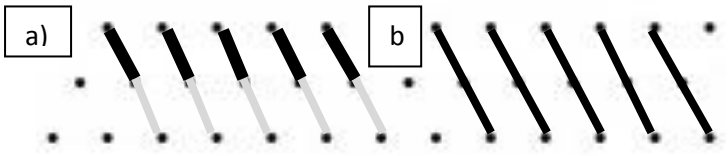


Figure 6.8: The merging of relations between, and nature of, Veps substance entities

Throughout history (and even now in the modern varieties, since we find new super- and subessive forms in the process of developing), Veps language-users have re-analysed where the right edge of the fully-inflected wordform is. That is, earlier in the language a substance entity (the postpositional phrase) consisted of two separate word-level substance entities **sugiden kerta* 'with the combs'. Here, there were two right edges: that at the end of the fully-inflected wordform *sugiden* and that at the end of the separate postposition **kerta*. The first line 'links' the genitive singular suffix *-n* to the lexical stem, and the second line 'links' the postposition **kerta* to the genitive form of the nominal.

The merging of the two 'separate' lines in figure 6.7 is seen in the increasingly fused relationship of the word-level substance entity and the postposition, gradually reanalysed as a case suffix.[28] Nevertheless, the form remains a substance entity - the level at which it 'sits' simply shifts gradually. To continue the metaphor of the visual example given in figure 6.7: the line joining the substance entities remains the same - since the forms encoded the same functions and were roughly phonologically equivalent (although phonetic erosion affected the final sound *-*rta* and the lexical stress of the postposition would have been lost) - but the way in which it is drawn alters diachronically.

Additionally, at both the word- and sub-word level, an invariant form can be associated with specific functions; this similarity of constancy in the function~form mapping aids in the development of a postposition such as **kerta* into a new morphological case-form.[29] Since Veps is already characterized by an abundance of sub-word substance entities, the introduction of more is in accordance with the other morphological structures found in the case system. That is, language-users are accustomed to the presence of relatively transparent sub-word structure; the introduction of novel case values as the result of suffixation of invariant phonological material in an isomorphic function~form mapping does not introduce any sort of unexpected or unusual morphological structure in the case system. In fact, I contend that it is more likely that case systems characterized by an abundance of sub-word substance entities than those with a lack of sub-word substance entities will develop new morphological case-forms from postpositional material. (The evidence from the exaptation of

[28] A very well-described grammaticalization pathway discussed by Blake (2001:163-169), Heine and Narrog (2011:516-518), and König (2011), for instance.

[29] It has been suggested (Kabak 2006) that the development of new morphological case-forms from postpositions is only possible where other morphological material such as case-forms do not intervene between the postposition and the nominal. Evidence from Turkish supports this hypothesis, but counter-evidence comes from various Finnic languages, including Veps. I suggest that the similarities between sub-word and word-level substance entities are more important in the reanalysis of postpositions as new morphological case-forms than whether or not a lexeme stem and postposition are 'interrupted' by inflexional material.

novel inflexional material in Romanian and Livonian may provide support for this hypothesis.)

The development of new morphological case-forms in Veps does not cause a shifting away from the reliance on sub-word substance entities, nor does it shift the locus of uncertainty away from that found in other inflected forms. Thus, the new inflexional case-forms exhibit similar morphological structure to that found throughout the Veps case system, and, I suggest, are therefore more likely to be 'accepted' by language-users, to spread through the system, and to persist diachronically.

New morphological case-forms thus tend to develop from postpositional material when a language is characterized by the following:

- an inflexional case system;
- an abundance of sub-word substance entities in the case system;
- postpositions (or other postpositional material such as adverbs and adverbial particles);
- a similarity of morphological structure at the sub-word- and word-levels.

Though the similarity of sub-word and word-level substance entities does not necessarily mean that the inflexional case system of a language will propagate morphological case-forms, as Finnish exhibits a lot of structure which is similar to that found in Veps, and yet has not expanded its case system in the same way.

6.5 CLOSING REMARKS

Drawing from the discussion offered in §§6.2-4, I present the following hypotheses:

- a language characterized (solely or predominantly) by schematic patterns is more likely to be case-reducing than those languages which are characterized by both schematic patterns and an abundance of sub-word substance entities;
- a language that is case-increasing is likely to be characterized by an abundance of sub-word substance entities;
- case-steady languages are of two types:
    - where there are ten or more morphological case-forms, the case system will be characterized predominantly by sub-word substance entities;[30]
    - where there are fewer than ten morphological case-forms, storage of whole wordforms and the employment of uncertainty-reducing schematic patterns (and sometimes paradigmatic schemata) provide predictive information not afforded by sub-word substance entities.

These hypotheses are summarized in table 6.2:

[30] The limit of ten morphological case-forms is very rough, and drawn from consideration of only the seven languages discussed in this chapter. However, it does accord with the generalizations offered by Plank (1986). I will attempt to make this estimate more accurate in chapter seven by interrogating a wider range of cross-linguistic data.

Case-reducing languages	Case-steady languages (<10 case values)	Case-steady languages (>10 case values)	Case-increasing languages
characterized largely by schematic patterns	characterized largely by schematic patterns	characterized largely by substance entities	characterized largely by substance entities
storage of word-level substance entities	storage of word-level substance entities	storage of word-level substance entities;	storage of word-level substance entities
--	--	(additional) storage of sub-word substance entities	(additional) storage of sub-word substance entities
implicational relations	strong implicational relations (sometimes correlated with extra-morphological considerations)	implicational relations	implicational relations
(likely) non-indexical case-markers	symbolic case-markers	indexical case-markers	symbolic case-markers

Table 6.2: Discussion from §§6.2-6.4 consolidated

Furthermore, understanding of the different loci identified and defined in §6.1 provides us with some insight into the ways in which the various languages develop:

	Locus of informativeness	Locus of uncertainty	Locus of predictiveness
Latin	fully-inflected wordform	suffix; sometimes the stem	implicational relations
Old French/old Occitan	fully-inflected word-form	suffix; sometimes the stem	implicational relations
Romanian	fully-inflected word-form	suffix and stem (whole wordform)	implicational relations; paradigm
Livonian	fully-inflected word-form; (some sub-word suffixes)	stem and suffix	implicational relations; (very few) suffixes
Finnish	fully-inflected word-form; sub-word suffixes	'link' between stem and suffix	implicational relations; suffixes
Veps	fully-inflected word-form; sub-word suffixes	'link' between stem and suffix	implicational relations; suffixes

Table 6.3: Different loci in the languages discussed in this study

Here, we see clearly that the presence of only one locus of predictiveness is found in those languages which have lost their case systems; where there is more than one locus of predictiveness, a language is more likely to be case-steady or case-increasing; multiple loci of informativeness provide a language-user with more information which may aid in the production of novel forms.

Furthermore, the development of new case-forms must match the structures already found elsewhere in the case system. For instance, since suffixal and postpositional material exhibit similar behaviour in Veps, it is more likely that postpositions will grammaticalize into novel case-forms through time. In Latin, on the other hand, new case-forms would have to be fused to the stem, cumulatively express case and number (and sometimes gender) values, and differ phonologically according to declension class. Thus, even if Latin had been characterized by postpositions, these were unlikely to have developed into new morphological case-forms, since they would have occurred with an already inflected form (and thus not have 'fused' with the stem in the same way as the other case-forms), would have expressed only case values (and not number or gender values), and would have been general across nominals, unlike any other inflexional case suffix in the language.

In Livonian and Romanian, the novel case-forms are not the result of the grammaticalization of postpositional material, but of the exaptation of morphophonological variation already found in the language. These novel forms, in a similar way to the new forms in Veps, match the language-specific morphological structures already found in the case system, which aids in language-user's acquisition and generalization of these exponents.

Finally, in old French in particular, the gradual erosion of discriminative morphology, the lack of indexical suffixal marking, and the preference for lexical invariance may have aided the generalization of similarly morphologically invariant adpositions. This similarity of

behaviour may have facilitated the reliance on minimally-entropic adpositions over the uncertainty associated with formal variation in the nominal in Latin, as well.

What I hope to have made clear in this chapter and in previous discussion is that no single factor can be identified which neatly explains the loss, maintenance, or accretion of inflexional case in a language. This, at least, I believe to be uncontroversial. Instead, in general, at least three main factors need to be taken into consideration when explaining changes that have affected morphological case systems:

- language-internal factors such as sound changes, analogical developments, functional fluctuation, and syntactic shifts;
- language-external pressures such as contact;
- the presence of substance entities and schematic patterns and the role(s) these play in the production of inflected case-forms.

The first two have been noted as important for the understanding of language change, either tacitly or explicitly, in earlier work. The third has been introduced here and provides, I suggest, a neat way of talking about the sorts of morphological structures found in language. However, the conclusions offered here are drawn from consideration of only seven languages, and only two language families. In order to reinforce the validity of these generalizations, it will be necessary to consider a wider range of languages. This will be done in a (necessarily brief) cross-linguistic survey in the next chapter.

7.

CROSS-LINGUISTIC SURVEY

7.0 INTRODUCTION

The observations from the discussion in chapter six were:

- a language characterized (solely or predominantly) by schematic patterns is more likely to be case-reducing than those languages which are characterized by both schematic patterns and an abundance of sub-word substance entities;
- a language that is case-increasing is likely to be characterized by an abundance of sub-word substance entities;
- case-steady languages are of two types:
    - where there are ten or more morphological case-forms, the case system will be characterized predominantly by sub-word substance entities;
    - where there are fewer than ten morphological case-forms, storage of whole wordforms and the employment of uncertainty-reducing schematic patterns (and sometimes paradigmatic schemata) provide predictive information not afforded by sub-word substance entities.

These hypotheses have to do largely with the ways in which inflexional case systems are likely to develop. These predictions apply to many other Indo-European and Uralic language families,[1] but further study is required to determine whether these observations hold more widely. Unfortunately, often the extant data are such that diachronic interrogation is not always possible, and I will not attempt this line of enquiry here. This chapter instead will be concerned with providing data to support *synchronic* predictions that can be drawn from the data already considered:

[1] Examples of Indo-European languages increasing their case systems (e.g., Tocharian, Lithuanian) and Uralic languages with 'smaller' case systems (northern Khanty, which has three case values) are rare.

- languages with around ten or more morphological case-forms are likely to be characterized by an abundance of sub-word substance entities (in addition to word-level substance entities and schematic patterns);
- languages with fewer than around ten morphological case-forms may be characterized by *either* a) an abundance of sub-word substance entities (in addition to word-level substance entities and schematic patterns), *or* b) schematic patterns (and word-level substance entities) alone.

If it is found that these hypotheses represent a cross-linguistic generalization, it may offer support for the diachronic predictions made in chapter six, too. After all, if it is not possible to find languages with both fusional morphological structure and more than ten case-forms synchronically, it might imply a limit on the number of case distinctions that can be maintained through time by schematic patterns alone. These synchronic observations should be supplemented by careful study of other cross-linguistic varieties in order to determine whether the diachronic generalizations apply more widely, although this will not be possible in this chapter.

7.1.0 THE CROSS-LINGUISTIC DATA

The preliminary conclusions drawn here from a handful of languages accord with those already offered, notably by Plank (1986), whose explanations for the limit on fusional structure in a case system with more than ten morphological case forms have to do with considerations of formal economy. Differentiating fusional from agglutinative morphology (the first of which is usually characterized by ‘sensitive cumulative’ markers, the second by

'insensitive separative' markers),[2] Plank argues that it is an 'arithmetical fact' that the latter is the most economical in the expression of multicategorial inflexional paradigms:

Insensitive separative:	[NOUN-NO.$_1$-CX$_1$]	[NOUN-NO.$_2$-CX$_1$]
	[NOUN-NO.$_1$-CX$_2$]	[NOUN-NO.$_2$-CX$_2$]
	[NOUN-NO.$_1$-CX$_3$]	[NOUN-NO.$_2$-CX$_3$]
	[NOUN-NO.$_1$-CX$_4$]	[NOUN-NO.$_2$-CX$_4$]

Table 7.1: Insensitive separative inflexional case-marking

In this example, language-users require knowledge of only six inflexional markers in order to correctly inflect a given lexeme. In a system characterized by insensitive cumulative forms, on the other hand, language-users must be familiar with eight distinct suffixes in order to correctly produce all forms in the inflexional paradigm; in a system characterized by sensitive separative elements, as many as ten different exponents must be known in the deduction of the correct inflected form:

Insensitive cumulative:	[NOUN-NO.$_1$CX$_1$]	[NOUN-NO.$_2$CX$_1$]
	[NOUN-NO.$_1$CX$_2$]	[NOUN-NO.$_2$CX$_2$]
	[NOUN-NO.$_1$CX$_3$]	[NOUN-NO.$_2$CX$_3$]
	[NOUN-NO.$_1$CX$_4$]	[NOUN-NO.$_2$CX$_4$]
Sensitive separative:	[NOUN-NO.$_1$-CX$_1$]	[NOUN-NO.$_2$-CX$_5$]
	[NOUN-NO.$_1$-CX$_2$]	[NOUN-NO.$_2$-CX$_6$]
	[NOUN-NO.$_1$-CX$_3$]	[NOUN-NO.$_2$-CX$_7$]
	[NOUN-NO.$_1$-CX$_4$]	[NOUN-NO.$_2$-CX$_8$]

Table 7.2: Insensitive cumulative and sensitive separative inflexional case-marking[3]

The view presented in this study is similar to that of Plank (1986) in that it is argued that considerations of morphological structure are central to understanding the implicational relationship of more agglutinative structure and languages with more than ten morphological case-forms. However, the hypotheses discussed here differ in that it considers formal

[2] Briefly: 'cumulative' markers are those which encode more than one feature value (i.e., case *and* number); 'separatist' markers mark distinct feature values linearly; 'sensitive' forms are distributed according to, e.g., number or wordclass, whilst 'insensitive' forms are found throughout the language regardless of number, declension class, etc..

[3] Tables 7.1 and 7.2 adapted from Plank (1986:41).

economy effects to be *emergent* properties of inflexional systems, reflecting an evolutionary pressure towards Low Conditional Entropy.[4] That is: "speakers' need to solve the [Paradigm Cell-Filling Problem] serves as a strong evolutionary pressure on language, guiding development of morphological systems in the direction predicted by the [Low Conditional Entropy Conjecture]."[5]

The work conducted here should therefore be considered a complementary addition to the typological studies of Plank (1986).[6] I have consulted the *World Atlas of Linguistic Structures* online (henceforth WALS) and have conducted research into the case systems of 161 languages listed as 'case' languages.[7] In this chapter I include discussion on languages characterized by 6-7, 8-9, and 10+ cases. Although the generalizations offered in chapter six hold across this data set, by more closely considering a wider range of languages, evidence may be found which indicates that the limit of fusional structure is actually lower, and that languages with more than eight or nine morphological case-forms are almost invariably characterized by predominantly sub-word substance entities.

There are a few necessary caveats that apply to the following discussion. In the first instance, although I have consulted nearly three hundred grammars of those varieties listed in WALS as 'case' languages, I will select only a couple of aspects of a couple of those with 6-7, 8-9, and 10+ case-forms to discuss in more detail here, since there is not enough space to describe each in depth. Furthermore, I will not attempt in this chapter to outline the diachronic developments of the case systems which are considered, not only because of limitations of space, but also because the data are often lacking. Determining whether these

[4] In a similar way, Ackerman and Malouf (2015) argue that the No Blur Principle of Carstairs-McCarthy (1994), a strict principle of formal economy in inflexional paradigms, is a 'special case' of the Low Conditional Entropy Conjecture.
[5] Ackerman and Malouf (2015:6).
[6] Also listed in the Universals Archive: https://typo.uni-konstanz.de/archive/nav/browse.php?number=1&PHPSESSID=o744a8ch38unuatqjpi40842e199dbq8. Accessed 11/05/2019, 11:15.
[7] 261 languages are included as 'case' languages, but 100 of these distinguish case only in pronouns and have therefore been omitted. The list is available online: https://wals.info/feature/49A#1/17/148.

varieties have developed in certain ways as the result of the morphological structures found in their case systems will have to be left to later research, data permitting. Only the synchronic correlation of the size of the case system and the presence or absence of sub-word substance entities will therefore be considered here.

It must also be noted that there are a number of limitations associated with relying on grammars of the 'case' languages of WALS. In the first place, I am aware that the grammars drawn on cannot hope to capture exhaustively the inflexional variation found in a language (accounting, for instance, for diatopic and diastratic variation). Moreover, many of the grammars consulted approach the language data from a constructivist perspective, often actually describing the case suffixes as 'morphemes'. Instances in which the form~function mapping may thus be obscured in favour of fore-fronting regularity, and, where certain phonologically-distinct suffixes are not distributed according to regular phonological criteria, this information may not be explicitly stated. I am necessarily reliant on the information gathered and presented by the scholars who have studied these languages, which means that my conclusions may be in some ways limited by the sources of information that are at my disposal.

Finally, in WALS, Blake's (1994:1) definition of case has been adopted, except that "a more general stance has been taken with regard to **clitic** and **adpositional** case marking."[8] Therefore, in addition to inflexional suffixes, clitic, particle, and adpositional forms are included as 'case' markers in the WALS data. I will discuss only those varieties which mark case inflexionally, although it is recognized that distinguishing morphological and syntactic case is frequently difficult. Notably, however, languages which have 'syntactic' expression of case do not alter my analysis, since these markers are generally invariant and may therefore be analysed as substance entities.

[8] https://wals.info/chapter/49. Accessed 06/09/2018, 14:13. Emphasis original.

Although the vast majority of the data presented in this chapter appear to support the (synchronic) generalizations offered above, it is important to note that the Samoyedic language Selkup, which is not listed in WALS but which will be discussed in §7.2, provides evidence against the hypotheses I have put forward. The conclusions reached here therefore do not constitute language universals. Indeed, I follow Evans and Levinson (2009) in believing that universal generalizations which hold absolutely cross-linguistically are vanishingly rare. Where recurrent patterns in natural language structures are found, these may be explained not as 'universals' that reflect some underlying organization of all human languages, but instead "as stable engineering solutions satisfying multiple design constraints, reflecting both cultural-historical factors and the constraints of human cognition."[9] These engineering solutions underlie the tendencies that have already been observed by, e.g., Greenberg (1963) and Plank (1986). Thus, although the generalizations found here do not constitute 'universals', they may indicate that a language with ≥8-9 case values not distinguished by means of minimally-entropic affixal material is harder for speakers to acquire and use. The result of this is that 'compensatory' factors which provide speakers with greater predictability are introduced into the case system. (For instance, E-complexity may be reduced, or greater sub-word clarity is achieved by the generalization of sub-word substance entities; or, indeed, a little of both - cf. Bertacca 2009 on the loss of the English case system.) Synchronically, therefore, the implicational relationship between morphological structure and the size of the inflexional case system reflects the cognitive abilities and limits of language-users cross-linguistically.

7.1.1 LANGUAGES WITH 6-7 MORPHOLOGICAL CASE-FORMS

According to WALS online, 37 of the 161 languages surveyed are characterized by six or seven case-forms. Of these, six do not fall under my analysis of 'case' languages as the

[9] Evans and Levinson (2009:429).

different case distinctions are encoded morphologically by means of clitics, particles or postpositions, or because of a lack of available data.[10] This leaves the 31 listed in table 7.3.

Language name	Language Family	Spoken in
Bengali (Chittagong)	Indo-European	Bangladesh
Chuvash	Turkic	Russia
Dyirbal	Pama-Nyungan	Australia
Georgian	Kartvelian	Georgia
Gimira/Bench(non)	Afro-Asiatic	Ethiopia
Huitoto (Minica)	Bora-Witoto	Columbia, Peru
Kalispel/Salish	Salishan	North central Montana
Kannada	Dravidian	India
Kanuri	Nilo-Saharan	Nigeria, Niger, Chad, Cameroon
Koasati	Muskogean	Elton, Louisiana
Krongo	Kadu	Sudan
Kunama	Nilo-Saharan	Eritrea, Ethiopia
Lithuanian	Indo-European	Lithuania
Malakmalak	Language isolate	Australia
Malayalam	Dravidian	India
Maricopa	Yuman	Maricopa County, Arizona
Meithei (Manipuri)	Sino-Tibetan	Northeast India, Bangladesh
Miwok (Southern Sierra)	Yok-Utian	California
Nenets	Uralic	Komi Republic
Nubian (Dongolese)	Nilo-Saharan	Nile River
Oromo (Harar)	Afro-Asiatic	Ethiopia, Kenya
Páez	Paezan	Colombia
Polish	Indo-European	Poland
Pomo (Southeastern)	Pomoan	Northern California
Romani (Kalderash)	Indo-European	Israel, Greece
Russian	Indo-European	Russia
Saami (Northern)	Uralic	Finland, Norway, Russia, Sweden
Shipibo-Konibo	Panoan	Peru, Brazil
Turkana	Nilo-Saharan	Kenya
Turkish	Turkic	Turkey
Yup'ik (Central)	Eskimo-Aleut	Western and southwestern Alaska

Table 7.3: Languages in WALS with 6-7 case-forms

Generally in this sample, we find a mixture of different morphological structures in the case system. Languages such as Turkish exhibit very agglutinative structure with many

[10] The languages omitted are: Aymara (central), Dani (Lower Grand Valley), Ika, Kewa, and Yawelmani.

instances of isomorphic sub-word form~function mapping; others, exhibit more fusional structure, where schematic patterns are the only means available to language-users in the production of novel forms (in particular: Russian, Oromo (Harar), Polish, and Turkana).

So far, therefore, the hypothesis that languages with fewer than ten productive morphological case-forms may be characterized by schematic patterns *and* sub-word substance entities or *just* schematic patterns holds.

7.1.2 LANGUAGES WITH 8-9 MORPHOLOGICAL CASE-FORMS

23 out of the 161 languages listed in WALS online are described as having eight or nine cases. Of these, 14 have been omitted from consideration here as the 'case-forms' are listed in consulted grammars as enclitics, particles, or postpositions, or because of a lack of available data.[11] This leaves the nine in table 7.4:

Language name	Language Family	Spoken in
Burushaski	Language isolate	Pakistan, India
Greenlandic (West)	Eskimo-Aleut	Greenland, Denmark
Hua (Yagaria)	Trans-New Guinea	Papua New Guinea
Ingush[12]	Nakh-Daghestanian	Republic of Ingushetia
Kala Lagaw Ya	Pama-Nyungan	Queensland
Nivich/Nivkh	Language isolate	Russia
Quechua (Imbabura)	Quechua II-B	Andes Mountains
Tlinglit*[13]	Dene-Yeniseian	Alaska, British Columbia, Yukon, Washington
Wambaya*	Mirndi	Australia
Yukaghir (Kolyma)	Uralic-Yukaghir	Russia

Table 7.4: Languages in WALS with 8-9 case-forms

[11] The languages omitted are: Alamblak, Arabela, Burmese, Coos (Hanis), Garo, Japanese, Khalkha, Mangarrayi, Mundari, Rama, Udihe, Ungarinjin, Wardaman, and Yidiny.

[12] WALS lists Ingush as a language with 10+ morphological case-forms; Nichols (2011:127) instead counts eight. For this reason, Ingush has been included here rather than in §7.1.3.

[13] Here and in table 7.7 in §7.1.3, an asterisk indicates that there are elements of the language which appear to prove problematic for the hypotheses under investigation and which will therefore be explored in more detail in the remainder of the section.

The case systems of most of these languages, according to the grammars consulted, are characterized predominantly by sub-word substance entities. There are, however, some instances in which the function~form mapping is interrupted in these languages and inflexional variation which might be better described as schematic patterns rather than sub-word substance entities. Intriguingly, however, these instances of more fusional structure appear to be 'supported' by uncertainty-reducing methods, such as will be discussed in reference to each of the varieties under consideration below.

In Tlingit, for instance, five of the so-called 'local' cases exhibit phonological variation across the language. This largely has to do with differences in tone, where the distribution of the different tones appears to be synchronically phonologically conditioned: the low tone (unmarked) follows when a nominal stem exhibits a high tone (marked by an acute accent) on final syllables, and the high tone occurs following a nominal stem exhibiting a low tone. This observation holds for the allative (*-déi*, *-dei*) and ablative (*-dáx*, *-dax*), although "[t]he perlative *-náx* sometimes shows alternate tone, but sometimes does not, and the reasons for this are not exactly clear, although it may be that the perlative is in transition from a postposition to a suffix."[14]

The distinct suffixal shapes of the instrumental also appear to be distributed according to the shape of the nominal stem immediately preceding. The form *-n* is found after open monosyllables, *-teen* after open (polysyllabic) syllables and *-een* elsewhere. Although "Robert Littlefield (p.c. 2007) suggests that [the instrumental] may be separating such that *-een ~ -n* is becoming comitative and *-teen* is becoming instrumental, ignoring the phonological environment."[15] In this way, the majority of the formal variation in the suffix found in the Tlingit case system is 'supported' in most instances by a minimally-entropic

[14] Crippen (2010:64).
[15] Crippen (2010:64).

relationship between suffixal shape and the phonological environment in which the case-marker occurs.

The only instance of synchronically non-transparent phonological variation in the suffix in Tlingit is found in the locative case, which can be expressed by means of a long final vowel, a final *-i*, or final *-Ø*. Unfortunately, Crippen (2010) does not offer any insight into the distribution of these forms - that is, whether they are distributed lexically or according to phonological, or grammatical (e.g., number), or purely morphological considerations. Further data would be needed to determine how language-users can predict the locative form of a previously unencountered nominal without having any minimally entropic sub-word material on which to rely.[16]

Tlingit appears to exhibit more fusional structure than the Finnic languages with more than ten inflexional case-forms. However, for all cases but the locative, it appears as though the suffixal exponents are a) transparently phonologically related, and b) distributed regularly according to phonological (or perhaps functional, i.e., the instrumental/comitative) considerations. It has already been argued, in reference to vowel harmony alternations in Finnish, that, where desinences are distributed according to regular and transparent phonological environments (which therefore provide predictive information about the shape of the suffixal material that follows a lexical stem), they still constitute examples of sub-word substance entities. Moreover, transparent and regularly-distributed suffixal shapes may be analysed as a 'compensatory' strategy, which enables the infrequent schematic patterns in the

[16] The most frequent semantic case-form in Finnish and Estonian - the illative - also exhibits non-phonologically-conditioned allomorphy and, in the Estonian 'old' illative form, more fusional structure. Although rare, therefore, we may conclude that semantic case-forms, as long as they are frequent, may exhibit some fusional structure. (N.B. also that the fusional illative forms in Estonian, as well as in Votic and Livonian, is the one being brought back 'into line' with the rest of the inflexional case system by the introduction of minimally-entropic sub-word material, *-sse* in Estonian, *-sõ* in Votic, and *-z* in Livonian; fn. 39 in §3.1.2.) This hypothesis cannot be tested with Tlingit, however, as Crippen (2010) does not offer any sort of indication of the relative frequency of the locative case.

Tlingit case system to be maintained by providing a speaker with sufficient predictive information about the distribution of different case elements.

Certain forms in Wambaya are also characterized by distinct shapes across number and//or declension, but, unlike in most instances in Tlingit, these distinct forms are not always easily described as being regularly distributed according to phonological environments. In the first instance, the ergative can be distinguished by four suffixal shapes: *-yi* with kinship nouns, *-ji* with the dual, *-nu* optionally following a *u*-final stem (this form is "in free variation with *-ni*"), and the form *-ni* which occurs in all other environments.[17] Although these suffixes are not necessarily transparently related (though some do exhibit similar structure, such as *-nV* or *-Ci*), they are regularly distributed through the language according to either wordclass or number (as with *-yi* and *-ji* respectively). As in Tlingit, the phonological variation exhibited by the Wambaya ergative case is correlated with extra-morphological considerations which may be used to 'support' the more fusional structure.

The genitive forms (*-nigan-, -nugan, -nagan-*) are transparently related and occur according to particular phonological environments, where the first vowel of the suffix is conditioned by the quality of the vowel with which the nominal stem ends. Similarly, the proprietive (*-uj-, -nguj*) and privative (*-aja, -yaja-, -waja-, -baja-*) desinences can be analysed as being a) synchronically transparently related, and b) regularly distributed according to the phonological environment in which they occur. They may therefore be considered examples of sub-word substance entities here.

There is, however, a potential 'opacifying' factor found in the genitive, proprietive, and privative suffixes (as well as the origin suffix, which has not yet been discussed), in that they inflect for gender (or inflexion class).

[17] Nordlinger (1998:81).

	Genitive	Origin	Proprietive	
			Absolutive	Non-absolutive[18]
Class I	*-niganji*	*-inji*	*-nguji*	*-ngunyi-*
Class II	*-niganya*	*-inya*	*-ngunya*	*-ngunya-*
Class III	--	--[19]	*-ngunyma*	*-ngunymi-*
Class IV	*-niganka*	*-inja*	*-nguja*	*-nguji-*

Table 7.5: Formal variation across Wambaya inflexion classes in the genitive, proprietive, and origin cases

Although there are some instances where the form~function mapping is opacified - for instance, the seemingly defective genitive and origin case-forms in class III nominals, the syncretism of the absolutive and non-absolutive class II proprietive forms, and the lack of nasal *-n-* in the class IV non-absolutive proprietive form *-nguji-* (where other non-absolutive forms are characterized by a nasal following the *-u-*) - formal variation across these case suffixes tends to be relatively transparent. The phonological relatedness of the privative suffix forms is even clearer and the phonological distribution even more transparent:

	***C*-final stem**	***u*-final stem**	***a*-final stem**[20]	***i*-final stem**
Class I	*-baji*	*-waji*	*-aji*	*-yaji*
Class II/III	*-bajama*	*-wajama*	*-ajama*	*-yajama*
Class IV	*-baja*	*-waja*	*-aja*	*-yaja*

Table 7.6: The absolutive gender forms of the Wambaya privative suffix

The distribution of these forms therefore appears to be regularly correlated with the phonological shape of the stem with which they occur.

Formally distinct suffixal material across number and/or declension appears to be more common in Wambaya than in Tlingit or in most of the other languages listed in table 7.6. Nevertheless, the exponent shapes are not as phonologically opaque as those found in Latin,

[18] 'Absolutive' case-forms are those suffixes which can occur at the end of the wordform; 'non-absolutive' must be followed by another case-form.

[19] Nordlinger (1988:93f.) does not offer an example of a class III genitive form, perhaps because "[m]ore commonly the dative suffix is used to mark possession [...] and there are thus only a limited number of examples of the genitive suffix in the corpus." The class IV genitive form is the most common in the language.

[20] "These forms are in free variation with *-waji, -wajana* and so on in this environment." Nordlinger (1998:98).

for instance. Intriguingly, the most extensive phonological variation are exhibited by the case suffixes of more 'semantic' case-forms such as the privative, origin, and proprietive, but this appears to be 'supported' by phonological transparency and by the predictive distribution of distinct case-forms, in a similar way as the majority of Tlingit formal variation.

The data considered here might indicate that the limit of schematic patterning is lower than that indicated by the Italic and Finnic evidence. That is, almost all of the languages listed in table 7.4 mark case either by means of word-level substance entities such as postpositions, clitics or particles, or by means of (largely) minimally entropic sub-word substance entities. Formally different suffixal shapes are generally restricted to grammatical case-forms or are distributed regularly in the language according (usually) to transparent phonological considerations. It is only really Livonian which exhibits eight case values and an abundance of fusional structure. Statistically, however, as long as those languages with more than ten morphological case-forms are also characterized by predominantly sub-word substance entities, we may suggest that the 'limit' of fusional structure in the inflexional case system is around eight or nine case values, rather than ten.

7.1.3 LANGUAGES WITH 10+ MORPHOLOGICAL CASE-FORMS

Finally, 24 of the 161 case languages given in WALS online are characterized by 10+ cases. Of these, nine express case by means of particles, enclitics and/or postpositions, or do not have sufficient data available; according to the grammars consulted, one has fewer than ten morphological case-forms (Ingush - §7.1.2).[21] This leaves fourteen languages in WALS which exhibit more than ten morphological case-forms:

[21] The languages omitted are: Awa Pit, Brahui, Epene Pedee, Gooniyandi, Hamtai, Ket, Nez Perce, Pitjantjatjara, and Toda.

Language name	Language Family	Spoken in	Number of cases
Basque	Language Isolate	Basque Country	11-12
Chukchi*	Chuktoko-Kamchatkan	Eastern Siberia	11-13
Estonian	Uralic	Estonia	11-14
Evenki	Tungusic	Russia and China	13
Finnish	Uralic	Finland	12-15
Hungarian	Uralic	Hungary	16-23
Hunzib	Daghestanian	Dagestan	12
Kayardild	Macro-Pama-Nyungan	South Wellesley Islands	23
Lak	Daghestanian	Central Daghestan	10
Lezgian*	Daghestanian	Russia, Azerbaijan	18
Martuthunira*	Pama-Nyungan	Western Australia	14-19
Mordvin (Erzya)	Uralic	Republic of Mordovia	9-12
Nunggubuyu	Macros-Gunwinyguan	Australia	11
Udmurt	Uralic	Republic of Udmurtia	15

Table 7.7: Languages in WALS with 10+ case-forms

The case systems of these varieties support the generalization given in §7.0 that those languages with more than ten morphological case-forms will exhibit an abundance of sub-word substance entities. However, there are some instances of more fusional structure in a couple of these languages which require closer consideration.

In Chukchi, "[t]he only morphological irregularities are in the marking of the absolutive singular. The absolutive singular markings are quite complex, with a mixture of lexical and morphophonological conditioning factors determining the appropriate form".[22] This sort of morphologically-determined formal variation is precisely the sort that is predicted not to occur in a language characterized by more than ten inflexional case-forms. I hypothesize that absolutive forms are likely to be amongst the most frequent in the language, and that the likelihood of storage of these inflected forms provides a language-user with the predictive

[22] Dunn (1999:103).

information not afforded by sub-word substance entities. However, without corpora counts, such a hypothesis remains speculative.

The remaining desinences are, according to Dunn's grammar of Chukchi, distributed according to phonological considerations. Some of these phonological environments are similar across different suffixes. For instance, the alternation of *-t-* and *-Ø-* in the forms of the comitative and the ergative occurs in the same environments: the ergative form *-te* and the comitative form *ɣe-___-te* (discussed below) follow a vowel, and the other exponents occur elsewhere. The locative, allative, and ablative desinences are also apparently distributed according to different phonological considerations.

More problematically, the Chukchi comitative, associative, and privative forms are encoded by means of a circumfix. According to the definition offered in §5.1.1, a circumfix does not constitute a 'sub-word substance entity' in the same way as an inflexional suffix does, since it was stressed that substance entities should be continuous and not interrupted by any other phonological material. It might be possible to argue that these sub-word forms are still relatively transparent and, although they are bound by slightly different patterns of distribution when compared to other forms (that is, one element must precede the lexical stem, the other must follow), these forms appear to be easy to segment away from fully-inflected wordforms. Since two of the three circumfixes are formally invariant, and the comitative is characterized by two circumfixal shapes (it is only the suffixal element which differs) which are distributed according to regular phonological environments, there is extra predictive power in the deduction of these case-forms, even though the circumfixes cannot be described as 'sub-word substance entities' in the same way as prefixes or suffixes.[23]

[23] An alternative analysis may be to suggest that the comitative, associative, and privative values in Chukchi are redundantly expressed by co-occurring prefixes and suffixes (after Corbin 1987, for instance). The nature of the evidence available makes it impossible to determine which of these analyses holds for the Chukchi data.

Lezgian exhibits a great deal of sub-word regularity and, for the most part, accords with my predictions. Only three forms need to be discussed in more detail: the ergative, inessive, and inelative. In the first place, the ergative is not characterized by suffixal marking. Instead, there are elements which can be added to the absolutive form and which create the stem with which all other inflexional suffixes occur; this bare stem form is found marking the ergative. The ergative is thus not characterized by any minimally-entropic sub-word form. However, since it is a grammatical case-form, and therefore likely to be frequently used and encountered in the language, Lezgian exhibits the same pattern as Chukchi, Finnish, and Veps, thus providing further evidence for the 'special' status of grammatical case-forms.

The inessive and inelative, however, as semantic case values, are expected to be expressed by sub-word substance entities. In the inessive and the inelative, the stressed final vowel of the stem changes to a distinct vowel: *-ú* > *-á*, *-ü* > *-é*, and *-í* > *-é*. In the inelative, a final *-j* follows this changed vowel except where the inessive form ends in *-e*, in which instance we find the inelative suffix *-äj*.[24] Intriguingly, here we not only find more fusional morphological structure, but also a paradigmatic consideration, where the form of the inelative depends not phonologically but implicatively on the shape of the inessive. This schematic pattern is restricted to a very small part of the inflexional case system, and may well be supported by 'paradigmatic' considerations, where knowledge of the inessive form will provide sufficient information about the shape of the inelative form, and vice versa.

Finally, Martuthunira exhibits a number of morphological aspects which may at first glance appear to undermine the hypotheses put forward here:[25]

[24] Haspelmath (1993:78).

[25] Importantly, Dench (1987:133) does not distinguish "a class of 'inflections' from a class of 'derivations'". In this way, the number of case feature values is likely to change depending on the definition of 'case' that a linguist follows. I include all the feature values identified by Dench which exhibit allomorphy for the sake of completeness.

	Form	Functions
Accusative	*-<u>ngu</u> ~ -ku ~ -yu ~ -V*	The direct object of a finite verb; temporal adjunct.
Genitive	*-<u>ngu</u> ~ -ku ~ -yu ~ -wu*	Possession
(Belonging)	*-<u>ngura</u> ~ -kura ~ -wura*	Denotes an entity which is controlled by another entity.
Effector	*-<u>ngku</u> ~ -lu ~ -tu ~ -rtu ~ -thu ~ -u*	Marks 'demoted' actors in passive clauses, second predicates of manner, and instrumental noun phrases in passive clauses.
Locative	*-<u>ngka</u> ~ -la ~ -ta ~ -rta ~ -tha ~ -a*	Marks spatial or temporal location.
Ablative	*- nguru*	'away from'
Allative	*-:rta ~ -nulyarra*	'to, towards'

Table 7.8: Case suffix exponents in the Martuthuniran case system[26]

The forms of the genitive and accusative cases are in many ways similar (and are supposed to have descended etymologically from a single case value).[27] In each instance, apart from the form found only with pronouns (which are therefore grammatically distributed), the different desinences appear to be transparently 'conditioned' by the phonological environment in which they occur:

Accusative	Genitive	Conditioning environments
-ku	*-ku*	Following a nasal
-yu	*-yu*	Following a lateral or rhotic
-V	*-wu*	Following a vowel

Table 7.9: The distribution of Martuthuniran accusative and genitive suffixal exponents

Since these forms are distributed according to the same phonological environment, it may also be said that they support one another paradigmatically. That is: a form with an accusative in *-ku* will also have a genitive in *-ku*, whilst an accusative form in *-V* will exhibit a genitive form in *-wu*.

[26] Underlined forms in table 7.8 are found only with pronouns and will therefore be omitted from discussion that follows.

[27] Dench (1987:134-136).

Similarly, three of the suffixal shapes associated with the allative value are found distributed according to regular phonological environments, where the form *-:rta* is *-arta* following *-a-*, *-irta* following *-i-*, and *-urta* following *-u-*.[28] The second allative form is actually semantically distinct: the form *-:rta* "focuses on the goal of the motion and essentially ignores the process by which participants in the clause arrived at this goal", whilst the form *-mulyarra* "focuses more on the motion itself".[29] These may therefore be better analysed as two distinct case values with two distinct markers, both of which are minimally-entropic across the language.

The 'belonging' case exponents are also distributed according to regular, generalized, and transparent phonological conditions. The form *-kura* occurs following a nasal whilst the form *-wura* appears elsewhere. It is less clear, however, what the conditioning factors for the distribution of the privative desinences are: "[t]he conditioning of the two forms is poorly understood at present though there are clear patterns. Most importantly, the **-wirraa** allomorph never precedes the accusative suffix, and is rare preceding the **-rru** 'NOW' clitic."[30]

Finally, the locative and effector exhibit very similar formal variation which, in a similar way to the genitive and accusative cases, can be analysed as being distributed according to phonological considerations, although these environments are arguably less clear than those listed above for the accusative and genitive case exponents:

[28] "So far I have not found an example of the direct allative on a consonant stem." Dench (1987:138).
[29] Dench (1987:167).
[30] Dench (1987:138f.). Emphasis original.

Locative	Effector	Conditioning environments
-la	*-lu*	Following stems with 3+ morae ending in a vowel
-ta	*-tu*	Following a stem ending in *-n-*
-rta	*-rtu*	Following a stem ending in *-rn-*
-tha	*-thu*	Following a stem ending in *-ny-* or *-nh-*
-a	*-u*	Following a stem ending in a rhotic or lateral

Table 7.10: The distribution of Martuthuniran locative and effector suffixal exponents

It is not possible to determine from the data considered whether either the locative or the effector cases are frequent enough case to 'support' so many distinct exponents (by being stored, for instance). What is most important, however, is that the locative 'moves in step' with the effector case. That is, where the effector form of a nominal ends in *-tu*, the locative case-form will end in *-ta* and so on, where the only difference is in the quality of the final vowel. In this way, therefore, although the formal variation by which the locative case is characterized can be considered to be 'supported' paradigmatically by the forms of the effector (or, indeed, vice versa, depending on frequency considerations).

Although there are a number of instances of formal variation in the suffix in the Martuthunira case system, therefore, it appears as though most of the distinct case-forms are distributed largely according to grammatical (word-class), phonological, semantic (as in the allative), or even paradigmatic considerations. This provides a multitude of often minimally-entropic means for the language-users to correctly generate the correct shape of a novel inflected forms, even where the sub-word form~function mapping is interrupted.

7.1.4 CLOSING REMARKS

Of the languages considered here, the vast majority appear to support the hypotheses put forward in §7.0. In fact, analysing the extra data from the previous sections, we may perhaps put forward more specific generalizations:

1. languages with more ≥8 case-forms are characterized by an abundance of sub-word substance entities (in addition to word-level substance entities and schematic patterns);[31]
2. languages with ≤8 case-forms can be characterized by both sub-word substance entities, word-level substance entities and schematic patterns, or just by word-level substance entities and schematic patterns;
3. where a language is characterized by a large case system, any fusional structure is almost entirely restricted to the grammatical sub-system; where fusional structure is found outside of the grammatical sub-system, it is likely to be 'supported' by compensatory factors (such as synchronically 'transparent' phonological variation in the stem or suffix, regular phonological or grammatical distribution generalizations, or 'paradigmatic' support from other inflected forms in the paradigm).

The evidence considered here seems to confirm that there is an implicational relationship between sub-word substance entities and 'large' (≥8 case-forms) case systems, and that this holds very generally cross-linguistically. Where forms are frequent and can be stored, implicational relations can be very powerful, since the patterns are encountered frequently enough to be well established in speakers' minds. However, where forms are less frequent, they are less likely to get stored and relations between fully-inflected forms are less likely to be as salient as they are between frequent forms (even if these frequent forms exhibit non-canonical structure). Greater predictiveness is provided by minimally-entropic sub-word substance entities, which is why they are so consistently found in those languages with ≥8 morphological case-forms.

[31] If Livonian is considered a serious enough exception, this may be changed to '≥9'.

7.2.0 EXCEPTIONS

Of the data considered, only one language presents significant counterevidence to the generalizations offered in §7.1.4. The Samoyedic language Selkup is characterized by more instances of fusional structure than any of the languages already discussed in this chapter, and yet maintains up to twelve morphological case-forms. In order to try to determine why this language challenges the generalizations that otherwise seem to hold over the data already considered, I will investigate Selkup in more detail in §7.2.1.

Furthermore, although the observations drawn from the Italic, Finnic, and cross-linguistic data seem to hold for a vast majority of case systems, the diachronic prediction that languages characterized solely by schematic patterns are likely to maintain smaller inflexional systems does not apply, *mutatis mutandis*, to observations about the verb cross-linguistically. That is, a verb system which is characterized solely by schematic patterns can exhibit great E-complexity, and does not necessarily lose feature values through time. This will be discussed further in §7.2.2 in reference to the development of the Italic verb.

7.2.1 THE SELKUP CASE SYSTEM

In this section I focus on Selkup, a Samoyedic language belonging to the Uralic family (and therefore distantly related to the Finnic languages) which has not been extensively studied. The data considered here are therefore drawn largely from Helimski (2006) and complemented with grammatical comments made in Helimski (1983). According to these sources, Selkup distinguishes twelve case values, four number values, as well as possessive and non-possessive values morphologically. Despite this extensive E-complexity, there are a number of fusional aspects in the Selkup case system which challenge the generalizations put forward in this chapter. In the first place, Selkup inflected nominal forms are characterized by two stem shapes - ‘first’ and ‘second’ - the former of which “serves as the main stem […] of

nouns", whilst "certain paradigmatic and derivational forms are based on its [the first stem's] modified variant".[32] Although generalizations about the morphophonological differences between the first and second stems can be offered,[33] they are not absolutely predictive.

Moreover, it is not just the shape of the stem that opacifies sub-word clarity in the inflexion of novel forms. There are also *sandhi* alternations between particularly the number markers and the case suffixes (for instance, between the plural exponent *-t* and the instrumental ending *-sä*); vowel alternations in certain 'purely' morphological, rather than phonological, formal variation in the case suffix; and the opaque ordering of sub-word elements. Each of these will be discussed here.

In the first place, the Selkup lexeme {NOM} 'God; heaven' is characterized by a number of fusional structures, including: allomorphy in the stem, mutliple distinct exponents for a single case value, and, in the possessive paradigms, different orderings of sub-word elements.

	Singular	**Dual**	**Plural**	**Collective**
Nominative	*nom*	*nopqɪ*	*nuut*	*nuulʲmï*
Accusative	*nuum*	*nopqɪm*	*nuutïm*	*nuulʲmïm*
Genitive	*nuun*	*nopqɪn*	*nuutïn*	*nuulʲmïn*
Instrumental	*nopsä*	*nopqɪsä*	*nuussä*	*nuulʲmïsä*
Caritive	*nomkåålïk*	*nopqɪkåålïk*	*nuutkåålïk*	*nuulʲmïkåålïk*
Translative	*nuutqo*	*nopqɪtqo*	*nuutïtqo*	*nuulʲmïtqo*
Co-ordinance	*nuuš^{ʲ}š^{ʲ}ak*	*nopqɪš^{ʲ}š^{ʲ}ak*	*nuutïš^{ʲ}š^{ʲ}ak*	*nuulʲmïš^{ʲ}š^{ʲ}ak*
Dative/allative	*nuunïk*	*nopqɪtkïnɪ*	*nuutïtkïnɪ*	*nuulʲmïnïk*
Illative	*nomtï*	*nopqɪtkïnɪ*	*nuutïtkïnɪ*	*nuulʲmïntï*
Locative	*nopqïn*	*nopqiɪqïn*	*nuutqïn*	*nuulʲmïïqïn*
Elative	*nopqïnï*	*nopqiɪqïnï*	*nuutqïnï*	*nuulʲmïïqïnï*
Prolative	*nommïn*	*nopqiɪmïn*	*nuunmïn*	*nuulʲmïïmïn*

Table 7.11: Selkup non-possessive forms of {NOM} 'God, heaven'[34]

[32] Helimski (2006:556).

[33] For instance: "If the first stem ends in a vowel, the two stems are identical. If the first stem ends in a consonant, the second stem acquires an additional final vowel (*ï*), or loses its final consonant (usually $\mathbf{N_3}$) [where a final nasal alternates with a homotopic stop or *-Ø*] and, in monosyllabic nouns, may contain a vowel quantitatively or qualitatively different from that of the first stem." Helimski (2006:557).

[34] Adapted from Helimski (2006:558f.).

Across the language, the most striking instance of fusional structure is that of allomorphy in the lexical stem. The Selkup lexeme {NOM} exhibits the relatively opaque stem alternation *nom ~ nuu-*, which may be described as 'irregular'.[35] Certainly, the alternations exhibited by other (usually polysyllabic) lexemes can be a great deal more transparent, where the difference between the first and the second stem is simply the addition of an element *-ï-* or the loss of the final consonant in the first stem.[36] However, even in these lexemes complex morphophonological alternations are found in certain inflected forms. In table 7.11, for instance, the final nasal *-m* alternates with the labial stop *-p-* when it is followed by *-q-*, and the plural marker *-t* assimilates to the *-s-* of the instrumental form and becomes *-n-* when followed by the nasal *-m-*. In Helimski's (2006) discussion on Selkup grammar, these alternations are described as instances of internal sandhi:

C1/C2	**k**	**t**	**q**	**s**	**š^{j}**	**m**	**n**	**l**
p						mm	mn	[βl]
t				ss	š^{j}š^{j}	nm	ŋn	ll
k			qq			ŋm	ŋn	[γl]
q	kk					qm	qn	ll
s					š^{j}š^{j}			
s^{j}				š^{j}š^{j}				
m								[βl]
n	ŋk							ll
n^{j}		n^{j}č^{j}					n^{j}n^{j}	
ŋ$_{1}$								[γl]
m$_{2,3}$	pk	pt	pq	ps	pš^{j}			[βl]
n$_{2,3}$	tk	tt	tq	ss	š^{j}š^{j}			ll
ŋ$_{2,3}$	kk	kt	qq	ks	kš^{j}			[γl]
l^{j}		l^{j}č^{j}						l^{j}l^{j}
j		j č^{j}						

Table 7.12: Sandhi in Selkup (where C1 and C2 belong to different 'morphemes')[37]

[35] Spencer (2013:396). Note that, although many of the other languages characterized by ≥8 morphological case-forms also exhibit allomorphy in the lexical stem, it is usually of a similar type to that exhibited by Finnish or Veps, in that morphological variations affect only the last sound of the stem. The morphophonological differences between particularly monosyllabic lexeme stems in Selkup are often more extensive.
[36] Although it does not appear to be predictable which of these a stem will exhibit.
[37] Adapted from Helimski (2006:556).

Of course, without further research it is impossible to determine whether these alternations are automatic for speakers of Selkup or whether they have morphologized in the modern language, although Helimski seems to analyse those in table 7.12 above as productive (as he lists separately those which "are no longer productive and regular").[38] There seems to be evidence to counter this suggestion, however. For instance, in table 7.12, there are different 'types' of nasals listed for C_1: *m*, *n*, *ŋ*$_{1,}$ *m*$_{2,}$ *m*$_3$, *n*$_2$, *n*$_3$, *ŋ*$_2$, and *ŋ*$_3$. Although *m*$_1$, *m*$_2$, and *m*$_3$ are all phonetically the 'same' sound, they are differentiated because of the different morphophonemic behaviour they exhibit:

> "Perhaps the most characteristic morphophonemic phenomenon of Selkup is the alternation of nasals with homotopic stops (*m* ~ *p*, *n* ~ *t*, *ŋ* ~ *k*), in some cases also with zero in word-final position, cf. […] *suurïm* ~ *suurïp* ~ *suurï* 'animal', *karman* ~ *karmat* 'pocket', *somaŋ* ~ *somak* 'well', *ååŋ* ~ *åå* 'mouth', etc. Such variants are functionally equal and they are almost equally frequent, for example, in pre-pausal position. The alternation is not automatic: word-final nasals and stops may be also stable, e.g. *këm* 'blood', *toop* 'edge' […]. Depending on the presence/absence and the conditions of this alternation, *several types of stems and suffixes can be distinguished*".[39]

Helimski identifies the following distinct ways in which forms inflect depending on the type of homotonic alternation of nasals (**N**) and stops (**T**) that they exhibit:

	N$_1$	**T$_1$**	**N$_2$**	**T$_2$**	**N$_3$**
Alternation in word-final position	*none*	*none*	**N~T**	**N~T**	**N~T~Ø**
Consonant typically preserved in inflexion	**N**	**T**	**N**	**T**	**Ø**

Table: 7.13: Morphophonemics in Selkup[40]

The behaviour exhibited by nominals in Selkup, as described by Helimski (2006), is not characteristic of phonological processes. A language-user requires knowledge of whether the lexeme they wish to inflect exhibits a particular homotonic alternation before they can

[38] 2006:555.
[39] Helimski (2006:554f.), emphasis added.
[40] Adapted from Helimski (2006:555).

determine which type of sandhi applies between the lexical stem and inflexional ending. It is less likely, therefore, that forms are as easily 'constructed' as in a language such as Turkish, where knowledge of an invariant lexical stem and invariant case suffix is frequently sufficient to produce the current inflected form of a nominal. Instead, Selkup appears in this respect more like Finnish, where implicational relations are required to determine the alternations that affect the 'link' between the lexical and the inflexional material, but where minimally-entropic case and number suffixes may be agglutinatively concatenated onto this form once it has been deduced.

In addition to the consonant alternations found in the forms in table 7.11, vowel alternations in Selkup nominal stems also appear to be frequent, as demonstrated by certain forms in the first person singular possessive paradigm of {NOM}:

	Singular	**Dual**	**Plural**	**Collective**
Nominative/ accusative	*nommï*	*nopqɪm(ï)*	*nuuiim(ï)*	*nuulʲmïmï*
Genitive	*nomnï ~ nuunï*	*nopqɪnï*	*nuuiinï*	*nuulʲmïnï*
Instrumental	*nomnïsä ~ nuunïsä*	*nopqɪnïsä*	*nuuiinïsä*	*nuulʲmïnïsä*
Caritive	*nomnïkåålïk ~ nuunïkåålïk*	*nopqɪnïkåålïk*	*nuuiinïkåålïk*	*nuulʲmïnïkåålïk*
Translative	*nomnoo(qo) ~ nuunoo(qo)*	*nopqɪnoo(qo)*	*nuuiinoo(qo)*	*nuulʲmïnoo(qo)*
Co-ordinance	*nomnïšʲak ~ nuunïšʲak*	*nopqɪnïšʲak*	*nuuiinïšʲak*	*nuulʲmïnïšʲak*
Dative/allative	*nomnïnïk ~ nuunïnïk*	*nopqɪnïkinɪ*	*nuuiinïkinɪ*	*nuulʲmïnïnïk*
Illative/locative/ elative	*nopqäk*	*nopqɪnïqäk*	*nuuiiqäk*	*nuulʲmïïqäk*
Prolative	*nommäk*	*nopqiɪmäk*	*nuuiimäk*	*nuulʲmïïmäk*

Table 7.14: 1st singular possessive forms of Selkup {NOM}

Of particular interest here are the translative forms across all numbers, where instead of the expected final *-ï-* of the possessive stem a long *-oo-* is found. Originally, this may have been an alternation caused by the following translative suffix *-qo*, but in the modern language

this element is optional, and so the alternation may be better described as a morphologized one.[41]

In addition to the stem, there are occasional instances in which case suffixes exhibit formal variation, which is not always as transparent as that found in the Finnic languages, nor always distributed according to clear phonological conditions, as in Chukchi. For instance, the dative/allative element in the singular and collective of the non-possessive forms (table 7.11) may be analysed as *-ïk*, but in the dual and plural we find a form *-tkïnɪ*; the prolative in the non-possessive and second/third person singular possessive forms may be analysed as *-mïn*, but in the first person singular possessive forms appears to be *-mäk*.

Finally, whilst in Finnish and Veps (and almost all inflected forms in those languages with more than eight morphological case-forms), feature values are linearly expressed and sub-word structure always follow the same syntagmatic order, in Selkup there are occasional instances where the ordering of sub-word elements does not follow that found elsewhere in the paradigm of a lexeme.

For instance, the element *-tï* may be analysed as the marker of the second/third person singular possessive,[42] and is found in all the forms in table 7.15 below (apart from the 2nd singular possessive forms and the translative, where *-tï* > *-too-* which was discussed above). However, where this form occurs in relation to the case marker can differ. In the instrumental, caritative, co-ordinative, and dative/allative, the possessive element precedes the case suffix; in the illative/locative/elative and prolative forms, on the other hand, the form *-tï* follows the case marker. Historically, this may reflect the development of new forms from

[41] Helimski (2006:557; 561) does not present these as changes occurring as the result of phonological environments, but rather as alternations that are conditioned by particular grammatical environments. In fact, the change *-ï-* > *-oo-* is not just the dual which licenses the alternation, but also before the translative (as already discussed), the illative, and the locative suffixes, a distribution that appears morphomic.

[42] Apart from in the nominative, where the 2nd and 3rd person possessive markers may be analysed as *-lï* and *-tï* respectively.

postpositions which would have followed the possessive form, but, synchronically, it offers another example in which the sub-word structure of the fully-inflected wordform is opacified which makes it more difficult for a language-user to produce the correct form by means of the concatenation of minimally-entropic sub-word substance entities.

		Singular	**Dual**	**Plural**	**Collective**
Nominative	2nd sing.	*nomlï*	*nopqɪlï*	*nuuiilï*	*nuulʲmïlï*
	3rd sing.	*nomtï*	*nopqɪtï*	*nuuiitï*	*nuulʲmïtï*
Accusative		*nomtï*	*nopqɪmtï*	*nuumtï*	*nuulʲmïmtï*
Genitive		*nomtï*	*nopqɪntï*	*nuuiintï*	*nuulʲmïntï*
Instrumental		*nomtïsä*	*nopqɪntïsä*	*nuuiintïsä*	*nuulʲmïntïsä*
Caritive		*nomtïkåålïk*	*nopqɪntïkåålïk*	*nuuiintïkåålïk*	*nuulʲmïntïkåålïk*
Translative		*nomtoo(qo)*	*nopqɪntoo(qo)*	*nuuiintoo(qo)*	*nuulʲmïntoo(qo)*
Co-ordinance		*nomtïšʲak*	*nopqɪlïšʲak*	*nuuiintïšʲak*	*nuulʲmïntïšʲak*
Dative/allative		*nomtïnïk*	*nopqɪɪtïnkinɪ*	*nuuiintïkinɪ*	*nuulʲmïntïnïk*
Illative/locative	2nd sing.	*nopqäntï*	*nopqɪɪqäntï*	*nuuiiqäntï*	*nuulʲmïïqäntï*
/elative	3rd sing.	*nopqïntï*	*nopqɪɪqïntï*	*nuuiiqïntï*	*nuulʲmïïqïntï*
Prolative	2nd sing.	*nommäntï*	*nopqɪɪmäntï*	*nuuiimäntï*	*nuulʲmïïmäntï*
	3rd sing.	*nommïntï*	*nopqɪɪmïntï*	*nuuiimïntï*	*nuulʲmïïmïntï*

Table 7.15: 2nd and 3rd singular possessive forms of Selkup {NOM}[43]

The relatively frequent instances of fusional structure in Selkup appear to violate the generalization that those languages with more than eight case feature values will be characterized by an abundance of sub-word substance entities. I will put forward a tentative hypothesis to explain this apparent counterexample in §7.2.3.

7.2.2 THE ROMANCE VERB AND ITS DEVELOPMENT

A brief consideration of certain verb systems shows that the diachronic generalizations given in §7.0 in reference to case systems do not necessarily apply to the development of the verb. That is: I have suggested that an abundance of sub-word substance entities can aid in the maintenance and propagation of case systems since they provide a language-user with additional, minimally-entropic means of deducing novel inflected forms, whilst fusional

[43] Selkup nominals also inflect for the dual and plural possessive in all three persons. There are no possessive forms for the collective.

morphological structure, in which the mapping of form and function is interrupted in various ways, often leads to the loss of the morphological distinction of certain feature values. In the Italic verb, however, there is a great deal of fusional structure, and yet the modern Romance verb maintains (and has even innovated) a great deal of E-complexity. This section will focus on briefly presenting the facts and offer some tentative reasons for the disparate developments of the noun and verb paradigms in the Italic languages.

I am certainly not the first to draw attention to the differences in the diachronic developments of the Latin nominals and the Latin verbs. Already, other scholars have noted the apparent disparity between the loss of inflexional categories in the nominal which the Romance verb has in many instances not only retained but has even propagated:

> "Tout d'abord, la flexion n'est pas traitée de la même façon dans le domaine nominal et dans le verbe. Dans les formes nominales, la flexion (déclinaison) est largement éliminée; dans les formes verbales, par contre, la flexion (conjugaison) est largement maintenue et même, en partie, refaite."[44]

> "The near-total elimination of an inflectional case system whose endings are characterized by extensive cumulativeness, syncretism and allomorphy (e.g., -O in DOMINO 'master' simultaneously signals dative and singular, is identical to the ablative singular case ending, and is restricted to second declension nouns, the same functions being performed by -I, -UI, etc., in other declensional classes), and its replacement in many Romance varieties by structures in which case is indicated configurationally or by means of prepositions, might suggest that the case system was lost, in part, because of sheer morphological complexity, but one need only point to the verb to find a system of even more complex inflectional morphology preserved, substantially, intact."[45]

In both of these quotations, it is stressed that the inflexional 'complexity' - both the E-complexity of the individual varieties and the ways in which these are encoded morphologically - of the modern Romance languages in respect of the verb is not simply that which has been inherited from Latin. Instead, there are also 'complexifying' developments

[44] Coşeriu (1987:58).
[45] Maiden (2011a:155f.). Emphasis added.

(e.g., the development of a new synthetic future form after the loss of Latin *-bo, -bis* etc.; the drastic increase in allomorphy in the verb stem).

Although I cannot offer an exhaustive account of the development of the Latin verb, I will give a brief overview of the 'basic' facts here; in §7.2.3, I will offer tentative explanations as to why we appear to find counterevidence in both the Selkup case system and in the Italic verb.

The Latin verb usually distinguished the following feature values morphologically:

- person (first, second, third);
- number (singular, plural);
- tense (present, future, imperfect, perfect, pluperfect, future perfect);
- voice (active, passive);
- mood (indicative, imperative, subjunctive);
- aspect (imperfective, perfective).

A typical exemplary paradigm might look something like that of {REGO} in table 7.16.[46] Here, the passive voice is only expressed morphologically in the indicative present, imperfect, and future; elsewhere, it is encoded by means of a periphrastic construction (the passive participle form of the verb in addition to a form of the verb {ESSE} 'to be' inflected in the relevant person/number, mood, and tense). There were thus analytic means of distinguishing certain morphosyntactic properties in Latin, just as there are in the modern Romance varieties. Furthermore, as in the Latin nominals, there is a great deal of fusional structure exhibited by the verb. It is not possible to separate suffixal elements into separate 'person' and 'number' markers,[47] and they exhibit formal differences not only across tense

[46] Vowel lengths have been omitted here since they will not be relevant to the discussion that follows.
[47] Matthews (1972).

(compare, for instance, the second person singular suffixes *-s*, *-bas*, *-isti*, *-(e)ris* etc.) but also voice. In addition, although restricted largely to a subset of third conjugation verbs, some lexemes are characterized by allomorphy in the stem (the stem shapes REG-, REX-, and RECT- in table 7.16). Whilst two of the stem allomorphs found in certain Latin verbs are correlated with imperfective and perfective aspect (REG- occurring with imperfective and REX- with perfective forms), the distribution of the stem RECT- is notoriously heterogeneous and cannot be aligned with any particular phonological or semantic considerations.[48] Furthermore, deriving one of these stem forms from another is difficult to do according to regular phonological processes that hold across the language.[49] Certainly, on even a brief consideration of the forms offered in table 7.16, there appear to be many instances of fusional structure and very few of unambiguous sub-word substance entities:

		Active		**Passive**	
Tense	**No.:**	**Singular**	**Plural**	**Singular**	**Plural**
		Indicative			
Pres.	**1st**	REGO	REGIMUS	REGOR	REGIMUR
	2nd	REGIS	REGITIS	REGERIS (-RE)	REGIMINI
	3rd	REGIT	REGUNT	REGITUR	REGUNTUR
Impf.	**1st**	REGEBAM	REGEBAMUS	REGEBAR	REGEBAMUR
	2nd	REGEBAS	REGEBATIS	REGEBARIS (-RE)	REGEBAMINI
	3rd	REGEBAT	REGEBANT	REGEBATUR	REGEBANTUR
Fut.	**1st**	REGAM	REGEMUS	REGAR	REGEMUR
	2nd	REGES	REGETIS	REGERIS (-RE)	REGEMINI
	3rd	REGET	REGENT	REGETUR	REGENTUR
Tense		**Indicative**			
Perf.	**1st**	REXI	REXIMUS	RECTUS (-A, -UM) SUM	RECTI (-AE, -A) SUMUS
	2nd	REXISTI	REXISTIS	RECTUS (-A, -UM) ES	RECTI (-AE, -A) ESTIS
	3rd	REXIT	REXERUNT (-ERE)	RECTUS (-A, -UM) EST	RECTI (-AE, -A) SUNT

[48] Aronoff (1994:31-59) and Maiden (2013) also offer extensive discussion on what they term the Latin 'third' stem.

[49] Matthews (1972) discusses at length the unsuitability of constructive accounts for the description of the Latin verb.

		Active		Passive	
Tense	**No.**	**Singular**	**Plural**	**Singular**	**Plural**
Pluperf.	**1st**	REXERAM	REXERAMUS	RECTUS (-A, -UM) ERAM	RECTI (-AE, -A) ERAMUS
	2nd	REXERAS	REXERATIS	RECTUS (-A, -UM) ERAS	RECTI (-AE, -A) ERATIS
	3rd	REXERAT	REXERANT	RECTUS (-A, -UM) ERAT	RECTI (-AE, -A) ERANT
Fut. perf.	**1st**	REXERO	REXERIMUS	RECTUS (-A, -UM) ERO	RECTI (-AE, -A) ERIMUS
	2nd	REXERIS	REXERITIS	RECTUS (-A, -UM) ERIS	RECTI (-AE, -A) ERITIS
	3rd	REXERIT	REXERINT	RECTUS (-A, -UM) ERIT	RECTI (-AE, -A) ERUNT
Pres.	**1st**	REGAM	REGAMUS	REGAR	REGAMUR
	2nd	REGAS	REGATIS	REGARIS (-RE)	REGAMINI
	3rd	REGAT	REGANT	REGATUR	REGANTUR
Impf.	**1st**	REGEREM	REGEREMUS	REGERER	REGEREMUR
	2nd	REGERES	REGERERIS	REGERERIS (-RE)	REGEREMINI
	3rd	REGERET	REGERENT	REGERETVR	REGERENTUR
Perf.	**1st**	REXERIM	REXERIMUS	RECTUS (-A, -UM) SIM	RECTI (-AE, -A) SIMUS
	2nd	REXERIS	REXERITIS	RECTUS (-A, -UM) SIS	RECTI (-AE, -A) SITIS
	3rd	REXERIT	REXERINT	RECTUS (-A, -UM) SIT	RECTI (-AE, -A) SINT
			Subjunctive		
Pluperf.	**1st**	REXISSEM	REXISSEMUS	RECTUS (-A, -UM) ESSEM	RECTI (-AE, -A) ESSEMUS
	2nd	REXISSES	REXISSETIS	RECTUS (-A, -VM) ESSES	RECTI (-AE, -A) ESSETIS
	3rd	REXISSET	REXISSENT	RECTUS (-A, -UM) ESSET	RECTI (-AE, -A) ESSENT
			Imperative		
Pres.	**2nd**	REGE	REGITE	---	---
			Infinitive		
Pres.		REGERE		REGI	
Perf.		REXISSE		RECTUS (-A, -UM) ESSE	
Fut.		RECTURUS (-A, -UM) ESSE		RECTUM IRI	

Table 7.16: Latin third conjugation {REGO} 'to rule, guide'[50]

For the Latin case system, it was suggested that this sort of morphological structure offered a partial reason for the loss of the morphological expression of case values through time; more generally it was argued that the presence solely of schematic patterns in a

[50] Participle, gerund, and supine forms are omitted.

morphological system was likely to lead to the loss of certain inflexional categories. But, as indicated by the comments offered by Coşeriu and Maiden above, the same arguments cannot be extended to the verb, where most of the feature values expressed synthetically in Latin continue to be expressed synthetically in the Romance languages. Most of the Romance languages distinguish the following feature values morphologically:

- person (first, second, third);
- number (singular, plural);
- tense;
- voice (active);
- mood (indicative, imperative, subjunctive, conditional);

Of course, a survey of the verb morphology of the Romance languages is beyond the scope of this study. The interested reader is referred to Lausberg (1973:259-380), Maiden (2011a, 2011b, 2018), Quer (2016), and Bertinetto and Squartini (2016) for further information. A short overview of some of the representative aspects of the Romance verb systems will be offered here.

In the first place, person/number markers from Latin are in many instances continued in the modern Romance languages. The first plural and second singular and plural suffixes *-mus*, *-s*, and *-(V)tis* respectively across tense and mood are continued very faithfully. For instance, the continuants of the first person plural marker *-mus* in selected Romance varieties in the paradigm of the Latin lexeme {CANTARE} ‘to sing’ are very faithful:

	Latin	Spanish	Italian
Pres.ind.	CANTAMUS	*cantamos*	*cantiamo*
Pres.sbjv.	CANTEMUS	*cantemos*	*cantiamo*
Impf.ind.	CANTABAMUUS	*cantábamos*	*cantavamo*
Perf.ind.	CANTAUIMUS	*cantamos*	*cantammo*

Table 7.17: Continuants of Latin first person plural marker *-mus*[51]

Despite phonological developments such as the loss of final *-m* (the 'default' first person singular marker), the remaining Latin person/number markers also tend to be retained in many of the Romance languages (although following the loss of final *-t* in the third person singular forms as well as that of final *-m* in the first person singular forms we find frequent instances in which these two sets of functions are encoded by the same form).

The passive of the present, imperfect, and future tenses synthetically in Latin is no longer expressed inflexionally in the Romance languages. This constitutes a clear instance in which we find a rise of configurationality similar to that of the increase in prepositional marking in the expression of case functions (discussed at length by Ledgeway 2012).

In terms of the number of tenses and moods expressed synthetically, the facts differ across Romance varieties. Portuguese has as many as nine or ten synthetic tense/mood distinctions: the present indicative, present subjunctive, imperfect indicative, imperfect subjunctive, preterite, pluperfect indicative, future, conditional, and future subjunctive (in addition to perhaps the inflected infinitive). Certain Romanian dialects spoken in central Transylvania, on the other hand, distinguish only the present and imperfect indicative morphologically.[52]

Many of the tense values expressed morphologically in the Romance languages are inherited from Latin. There are a handful of notable exceptions. In the first place, the perfect in the Romance languages is generally expressed by means of an auxiliary (a form of the verb

[51] Adapted from Maiden (2016b:498).
[52] Maiden (2016b:497).

'to have') and a participial form of the lexical verb,[53] whilst in Latin, perfect functions were expressed by means of a marker *-v/u-* (or, in certain third conjugation verbs, by allomorphy in the stem) which occurred before the person/number ending. The loss of this inflexional tense marking accords with the prediction that, where a system is characterized by schematic patterns alone, it should decrease the number of feature values it distinguishes morphologically.

The introduction of the new future tense form, however, does not accord with the predictions put forward at the beginning of the chapter. Whilst there was a synthetic future in Latin, it is not continued into Romance.[54] Instead, French and Italian, for instance, have developed an innovative future form from a periphrastic construction with the Latin verb {HABERE} 'to have', e.g., CANTARE HABEO > **cantarábeo* > **cantaráḭo*.[55] Thus we find the following forms in French and Italian:

	French		**Italian**	
	Singular	**Plural**	**Singular**	**Plural**
1st	*chanterai*	*chanterons*	*canterò*	*canteremo*
2nd	*chanteras*	*chanterez*	*canterai*	*canterete*
3rd	*chantera*	*chanteront*	*canterà*	*canteranno*

Table 7.18: Future forms in French {CHANTER} and Italian {CANTARE} 'to sing'

Finally, one of the most striking instances of 'complexification' in the Romance verb is that of the increase of allomorphy in the lexical stem. As already mentioned, in Latin, stem allomorphy is mostly found in certain third declension lexemes, but is the exception rather than the rule. A maximum of three distinct stem shapes are found, two of which are distributed according to aspect, whilst "allomorphy correlated with person, number and tense was nearly absent in Latin, and the rise of such allomorphy is a major innovation in

[53] Maiden (2011a:155).
[54] Lausberg (1973:310).
[55] Lausberg (1973:315).

Romance."[56] Some of this formal variation is inherited from Latin, but many instances are the result of phonological developments (and subsequent analogical spread throughout the language).[57] Although this development does not add to E-complexity of the Italic verb, it is an example of an increase in the fusional structure found in the Romance languages. This increased fusional structure, the introduction of novel inflexional forms, and the retention of much of the original E-complexity of the Latin verb all constitute counterexamples to the generalizations offered in §7.0.

7.2.3 ANALYSIS

The evidence from the Selkup case system and the Italic verb seems to present some difficulties for my hypothesis. In the first place, Selkup maintains a case system with more than ten morphological case-forms, and yet is also characterized by more instances of fusional structure (which is not just restricted to the grammatical case-forms) than are found in other languages with similarly large case systems. In addition, not only does the Romance verb maintain much of the original E-complexity of the Latin verb (as well as the fusional morphological structure), but further instances in which the form~function mapping is disturbed have been introduced diachronically, as is seen particularly in the startling increase in allomorphy in the stem.

I offer here three suggestions to explain the different developments found in the Italic noun and the verb, as well as for the large case system of Selkup:

- the E-complexity maintained by a given case (or verb) system;
- the size of the morphological family;

[56] Maiden (2011b:218).
[57] The phonological development of allomorphy in the lexical stem and its analogical spread throughout the various Romance languages is discussed by Maiden (2018).

- the nature of feature values marked inflexionally (i.e., whether they are more or less 'relevant' - Bybee 1985).

Each of these will be discussed in turn in reference to the data outlined in §§7.2.1-2.

First, compared to the languages already discussed, Selkup exhibits greater E-complexity in its case system. In most of the languages listed in WALS with 8-9 or 10+ case-forms, nominals regularly inflect for number and case; gender distinctions are relatively rare, as is inflexional marking of other feature values such as possession or definiteness. Moreover, whilst there may more than ten case values, number values rarely exceed two. In this way, the E-complexity of those languages characterized by more than eight morphological case-forms is relatively low. Selkup, however, exhibits not only twelve case values, but also four number values *and* possessive paradigms for the first, second, and third persons in all numbers but the collective. This means that the inflexional paradigm of a given nominal in Selkup has 480 cells, and thus exhibits greater E-complexity than the other languages discussed in this study.[58]

This increased E-complexity means that there are many more forms in the paradigm of a given lexeme which may need to be deduced by a speaker. Since storage of fully-inflected wordforms decreases in likelihood with the increase in the size of the inflexional paradigm, alternative means are required in the production of previously-unencountered forms. With such frequent instances of disturbance in the form~function mapping at a sub-word level, however, what besides the storage of minimally-entropic forms can provide sufficient information to 'support' such a large case system?

[58] cf. Finnish, for instance, which exhibits a far larger inflexional case system than Romanian or even than Latin, and yet which still only has a paradigm with up to thirty cells (counting two number values and the upper counter of fifteen case values). If the possessive forms *-ni, -si* etc. are included (analysed as suffixes rather than clitics after, e.g., Kanerva (1987), then the number of cells increases to 180. The 'support' provided by a lexeme's morphological family will be discussed later in this section.

One means might be the regular distribution of the sub-word recurrent partials of the different stem shapes that occur within the inflexional paradigm of a nominal. In this way, the case system of Selkup bears some similarities to the related Samoyedic variety Tundra Nenets, as discussed by Ackerman et al. (2009), a language which distinguishes eight case values, three number values, and three possessor values in three numbers. A regular Tundra Nenets nominal therefore has 210 cells in its inflexional paradigm, and is thus under half the size of the Selkup nominal paradigm.

In Ackerman et al.'s (2009:72-80) discussion of Tundra Nenets, the inflexional case system is analysed as forming 'alliances' between forms characterized by the same recurrent partials - in this instance, the shape of the lexical stem. The lexeme {NGANO} 'boat' has three recurring stems: *ngano*, *nganu*, and *nganoxo*; all of the inflected forms in this lexeme's paradigm are characterized by one of these stem shapes. These three forms participate in "a system of interpredictability among related word forms".[59] Although the different stem shapes cannot be 'derived' or predicted from one another with absolute certainty, where forms characterized by these distinct forms are known, and where a language-user is familiar with the distribution of the distinct stem shapes in the inflexional paradigm, this information can be utilized in the deduction of novel inflected forms.

Similar observations may be made for Selkup. Although the two stems of a lexeme such as {NOM} are not predictable from one another, once a language-user knows the two shapes, they may utilize implicative information about the distribution of these forms in the production of all other inflected items in the lexeme's paradigm. The recurrent partials *noC-* (where the final consonant alternates depending on the sounds that follow) and *nuu-* are regularly distributed in the way outlined in table 7.19.

[59] Ackerman et al. (2009:74).

	noC-	***nuu-***
Non-possessive	singular: nominative, instrumental, caritive, illative, locative, prolative; dual: all case-forms.	singular: genitive, accusative, translative, co-ordinative, dative/allative; plural: all case-forms; collective: all case-forms.
1st possessive singular	singular: all case-forms; dual: all case-forms	singular: genitive, accusative, translative, co-ordinative, dative/allative; plural: all case-forms; collective: all case-forms.
2nd/3rd possessive singular	singular: all case-forms; dual: all case-forms	plural: all case-forms; collective: all case-forms
1st possessive dual	singular: all case-forms; dual: all case-forms	singular: genitive, accusative, translative, co-ordinative, dative/allative; plural: all case-forms; collective: all case-forms.
2nd/3rd possessive dual	singular: all case-forms; dual: all case-forms	plural: all case-forms; collective: all case-forms
1st possessive plural	singular: all case-forms; dual: all case-forms	singular: genitive, accusative, translative, co-ordinative, dative/allative; plural: all case-forms; collective: all case-forms.
2nd/3rd possessive plural	singular: all case-forms; dual: all case-forms	plural: all case-forms; collective: all case-forms

Table 7.19: The distribution of distinct stem allomorphs in Selkup[60]

This morphomic distribution of stem shapes appears to be regular across the language, which lends it predictive power in the inflexion of novel wordforms. These patterned occurrences of recurrent partials constitute alliances of forms: where an item which is characterized by the recurrent partial *nuu-* or *noC-*, it provides information about the other forms characterized by these stem shapes. Even if, therefore, a dual form has not been encountered before, the relevant forms of the singular, or any form in the plural of any possessive or non-possessive paradigm, will provide sufficient information about the shape of the stem, whilst knowledge of the dual marker *-qı* and of the case markers (taken from other

[60] There is some overlap in the distribution of these two stems. In the singular of all first person possessive forms (in the singular, dual, and plural), there are two forms for the genitive, instrumental, caritive, translative, co-ordinative, and dative/allative, each based on one of the two distinct stems. Those based on the *nuu-* stem follow the distribution found in the singular non-possessive paradigm, whilst those based on the *noC-* stem follow the distribution of stem forms found in the second/third possessive paradigms.

dual forms known to the language-user), will enable all dual forms of a lexeme such as {NOM} to be derived.[61]

In addition to greater E-complexity in the inflexional paradigm, the morphological family of any given Selkup lexeme appears to be extremely large (according to the work conducted by Helimski 2006). I suggest that the larger the size of the morphological family of a lexeme, the more information about a novel form may be provided by forms outside of the inflexional case paradigm, which may help to reduce uncertainty in the production of a novel inflected form. The hypothesis that the size and structure of the morphological family of a lexeme has an effect on the processing of inflected forms is supported by evidence from psycholinguistics.[62] I suggest that it is also likely to aid with the production of previously-unencountered forms.[63]

In particular, the complex allomorphy of the stem in Selkup nominals may be 'supported' by a large morphological family (in addition to the extensive inflexional case system). For instance, 'relative' adjectival forms end in *-L*^{j} and are characterized by the second stem of a nominal, e.g., *nom* nominative singular 'god, heaven', relative adjective *nuul*^{j} 'godlike, heavenly'.[64] This form - or, indeed, any other given related form in the morphological family - may be exploited in the production of novel forms and may, overall, help to reduce the uncertainty associated with deducing a target form. A (very simple) example is offered in table 7.20:

[61] The "[s]ystematic mapping of root allomorphs onto an arbitrary set of paradigm cells" is found throughout the Romance verb system, too, and appears similarly to offer predictive information in the production of novel forms. Maiden (2018:303 *et passim*).
[62] Schreuder and Baayen (1997), de Jong et al. (2000), del Prado Martín et al. (2004), amongst others.
[63] A position also espoused by Bonami and Strnadová (2018).
[64] Helimski (2006:562).

(Non-possessed)	Singular	Dual	Singular	Dual
Nominative	*nom*	--	--	--
Genitive	*nuun*	*nopqɪn*	**?**	*qoqqɪn*
Translative	*nuutqo*	--	--	*qoqqɪtko*
	'Derived' forms			
Relative adjective	*nuul*^{j}		*qool*^{j}	

Table 7.20: Deduction of the genitive singular form of {QOŋ} 'chief'

The deduction of the genitive singular form of the lexeme {QOŋ} here involves relatively complex analogical deduction - more complex than the usual four-part analogical pattern - which is intended to reflect the probability that a language-user will not have all access to all the necessary information to be able to form a proportional analogy. Here, relevant information about the deduction of the genitive singular form of the lexeme {QOŋ} may be gleaned from both the inflexional paradigm and the wider morphological family. Let us say that a language-user is familiar with only the forms given in table 7.20 (reflecting the likelihood that a speaker is unlikely to have encountered all inflected forms of a given lexeme).[65] However, with reference to (at least) four different implication patterns that exist across certain morphologically-related forms in the language, a speaker may correctly deduce the shape of the genitive singular of {QOŋ}:

1. where a nominal is characterized by two distinct stem shapes, the genitive singular is characterized by the second stem;
2. the second stem is also found in the relative adjective form;
3. the relative adjective form is distinguished by means of a final -L^{j};
4. the genitive marker is -*n*;

No inflected form given in table 7.20 for the lexeme {QOŋ} is characterized by the second stem, whilst the 'derived' form of the relative adjective is. Pattern three allows a language-user to determine that genitive singular form is therefore characterized by the same stem

[65] Zipf (1935, 1949).

shape (distinct from the first stem shape found in, for instance, the nominative singular form) as the relative adjective. Knowledge of pattern four permits a language-user to remove the final -*lʲ* of the form *qoolʲ* to leave the bare second stem, and knowledge of pattern five permits a speaker to determine that the ending -*n* will discriminate genitive (singular) functions from other case values. The larger the morphological family (as well as the inflexional paradigm), the more likely it is that a language-user will have knowledge of the two stem shapes of a given lexeme.

The importance of the morphological family in supporting a large case system may also be seen in the Finnic data already discussed. Finnish nominals, for instance, frequently exhibit very large morphological families. Nominals can exhibit nearly 2,000 (productive) distinct inflected forms in their inflexional paradigm;[66] when derivational forms are included, this number can increase drastically, depending on the productivity of the lexeme. For example, {TYÖ} 'work' has around 7,000 family members (not to mention the inflected forms of these items).[67] The largeness of the morphological family in Finnish leads to shorter response latencies in processing,[68] and, as previously argued, it is also possible that larger language family provide predictive information for the speaker which may be utilized in the production of novel inflected (or, indeed, derived) forms.

In the Italic languages, on the other hand, although the morphological family can be large (depending, again, on the lexeme), there are far fewer inflected forms of family members than in the Finnic languages. There are therefore fewer forms in a nominal's morphological family in the Romance languages, and therefore it is possible that, on average, a language-user is exposed to less relevant information about the inflexional patterns exhibited by a

[66] Karlsson (1982:356).
[67] del Prado Martín et al. (2004:1272).
[68] del Prado Martín et al. (2004:1271).

given lexeme in the Italic languages when compared to the amount of information known to a speaker of a Finnic variety.

In the Romance verb, the production of novel forms may be similarly 'supported', not by the presence of an abundance of sub-word substance entities, but by large inflexional paradigms. In the first place, a similar analysis outlined in reference to Tundra Nenets and Selkup above may be applied to the Latin verb system, wherein the three stem shapes of a lexeme such as {REGO} may be analysed as forming systematic formal alliances:

REG-	**REX-**	**RECT-**
[-perfective]	[+perfective]	Past passive participle
-	-	Future active participle
-	-	Supine

Table 7.21: The paradigmatic distribution of distinct Latin verb stem allomorphs

The abstract relations that exist between known forms and which provide information about the distribution of these different recurrent partials may be utilized in the deduction of novel inflected forms. Since different stem shapes in the verb stem are relatively rare, this sort of distribution may help to support the allomorphy that cannot always be predicted according to phonological or, in the case of the *rect-* (or 'third') stem, grammatical considerations. It is likely that the language-user will meet sufficient forms of a given lexical item which will reduce uncertainty about the shape of all different stems in the lexeme's paradigm, or that that they will have information about the stem alternations found in verbs in the lexical neighbourhood.

The morphological family will also help to support this type of fusional structure. For instance, in Latin, distinct verb stems are not just confined to the inflexional paradigm but are found in certain derived forms, too. Latin agent nouns in *-or* exhibit a stem shape the same as that found in the perfect passive participle, e.g., SCRIPTOR 'writer' from SCRIB-O, SCRIP-SI,

SCRIPT-UM ‘to write’; FACTOR ‘maker’ from FACI-O, FEC-I, FACT-UM ‘to do, make’.[69] These agent nouns provide information about the shape of the stem that will be found in certain cells in the inflexional paradigm of the related verb. Thus, these sub-word recurrent partials partake in alliances of forms in a similar way to the stem allomorphs found in the Selkup (and Tundra Nenets) nominal paradigms.

Recognition of the informativeness not only of other inflected forms of a lexeme but also of derived forms accords with the predictions offered in §1.2.1 about the ‘opportunistic’ means employed by language-users in the production (and interpretation) of novel forms. I argued that it was possible to utilize any and all morphological patterns that could be recognized by language-users in the deduction of previously-unencountered inflected forms. If there exists, in Selkup or in Latin, for instance, a consistent relationship between form and function in the inflexional alternations exhibited by particular forms (which are morphologically related), then it is predicted that these may be extended in the production of novel forms, even if a traditional grammar analyses one as an ‘inflexional’ and the other as a ‘derived’ form:

SCRIPTOR : SCRIPTUS

[agent noun] : [perfect passive participle]

FACTOR : X

X : FACTUS

Figure 7.1: The analogical deduction of unknown Latin past passive participle forms

[69] cf. Steriade (2016), who claims that the distribution of Latin ‘stem syncretism’ is conditioned by the phonological shape of the suffix with which it occurs.

It does not matter that the forms FACTOR and SCRIPTOR are agent nouns (and may be described as 'derived' forms from the verbs {FACERE} and {SCRIBERE} respectively) and that FACTUS and SCRIPTUS are participial forms (which are usually analysed as part of the passive inflexional paradigm). Since these two forms are consistently characterized by the same recurrent partial, such a morphological pattern may be utilized in the deduction of a novel, e.g., perfect participle form (as in figure 7.1 above). If my suggestion that language-users make use of all and any patterns that provide them with information pertaining to the production of novel forms, then it is likely that, in addition to inflected forms of a lexeme, language-users will also use the morphologically-related items of a nominal in order to solve the Paradigm Cell-Filling Problem.

My third and final suggestion to account for the counterexamples identified in this chapter is that there are essential differences in the *nature* of the noun and the verb which have not, heretofore, been discussed. These differences have to do with what Bybee (1985) terms 'relevance', what Coşeriu (1987) calls 'déterminations internes' and 'déterminations externes', and what Booij (1994) calls 'inherent' and 'contextual' inflection. According to Bybee, "[i]f two meaning elements are, by their content, highly relevant to one another, then it is predicted that they may have lexical or inflectional expression, but if they are irrelevant to one another, then their combination will be restricted to syntactic expression."[70] For instance, aspect and tense marking in particular are very relevant to the meaning of the verb, since they directly modify the semantic content of the lexeme they mark. Cross-linguistically these categories, Bybee predicts, are more likely to be expressed synthetically (or by means of different lexemes), rather than analytically.

Similarly, Booij (1994, 1996) differentiates 'inherent' from 'contextual' information. The former refers to inflexion not required by the syntactic context (including number in nouns

[70] 1985:13.

and degree in adjectives), whilst the latter refers to inflexion which *is* required by the syntax (including grammatical case for noun and person/number markers on verbs). Booij argues that inherent information (that which is 'more relevant' to the meaning of the lexeme) tends to be marked closer to the lexeme stem than contextual information (that which is more relevant to the syntax), and that it is more likely to be expressed synthetically than analytically.

For Coşeriu (1987), those features which are more 'relevant' or 'inherent' to the nominal or verb are described as 'déterminations internes' whilst those which have less relevance (and are therefore 'contextual') are described as 'déterminations externes'. He uses these distinctions to help explain the loss of the Latin case system:

> "Ainsi, si, pour les fonctions des cas, la flexion - c'est-à-dire la détermination "interne" - est remplacée par l'expression périphrastique, c'est-à-dire par une détermination syntagmatique ou externe, c'est parce que les cas représentent des fonctions typiquement relationnelles: ils signifient un rapport entre le terme déterminée et un autre terme (ou d'autres termes) de la phrase. Par contre, pour le nombre et le genre, fonctions non-relationnelles c'est la détermination matérielle paradigmatique ou interne qui s'affirme."[71]

Thus the prediction is that number marking is likely to be more relevant, particularly for count lexemes, to the meaning of the nominal; by extension, it is hypothesized that number values will be marked morphologically, which is indeed what is found throughout Romance.[72] On the other hand, those functions which are less relevant to the meaning of the nominal are likely to be those which are expressed analytically rather than synthetically, which matches the rise of configurationality and the increase in prepositional marking for the marking of case in the Italic languages. The development of periphrastic marking for case whilst number remains expressed morphologically may be at least partly explained by the distinction between 'inherent' (or more relevant) and 'contextual (or less relevant). However,

[71] Coşeriu (1987:61).
[72] Maiden (2016c and 2011a:163 -166).

it should be noted that this tendency is by no means absolute, as shown by the developments particularly of Finnish and Veps (and perhaps also Romanian), and should be considered alongside other factors in the development of an inflexional case or verb system.

It is thus argued that a combination of extensive E-complexity, large morphological families, and the preference for the morphological marking of 'more relevant' or 'inherent' categories over 'less relevant' or 'contextual' categories may help to account for the Selkup inflexional case system and the development of the Italic verb system, both of which offer counterevidence to the generalizations offered at the beginning of this chapter. The existence of extensive E-complexity means that a lexeme is characterized by a large inflexional paradigm, which may mean that language-users are more likely to encounter at least one form of a given lexeme; this then provides some (if not a sufficient amount of) information about the shape of other forms in the inflexional paradigm. Moreover, if there is a large morphological family in addition, then there is likely to be even more information which may be utilized in the production of novel forms. Finally, the cross-linguistic preference for marking relevant (inherent) categories synthetically and less relevant (contextual) categories analytically is reflected in the diachronic development of the Italic nominal, and may offer partial reason for the continued morphological expression of number values across Romance whilst inflexional case is almost entirely lost.

7.3 CLOSING REMARKS

Whilst the data set considered from WALS appears to provide support for the synchronic generalizations discussed in this study, the evidence from Selkup and the development of the Latin verb seems to constitute counter-evidence for the observations drawn from the Italic and Finnic languages in particular. It has been suggested that E-complexity and the size of the morphological family as 'supporting' aspects of an inflexional system, providing language-

users with the predictive information that is sacrificed by the (more discriminative) fusional structures in the inflexional paradigm of the Selkup nominal and the Italic verb. Moreover, the near-total loss of the category of case is rendered less surprising when it is recognized that the category is less relevant to the meaning of a nominal than, for instance, the category of number, and therefore more likely, cross-linguistically, to be expressed analytically rather than synthetically.

It should be noted that the explanations offered here are preliminary in nature; more research is required to determine precisely why, for instance, the fusional morphology of the Italic verb has been so robustly maintained through time (particularly when compared to the Italic nominal). Additionally, although previous psycholinguistic studies have indicated that larger morphological families play a facilitating role in the *processing* of lexemes, there remains further work to be conducted on how exactly the size of the morphological family aids in the *production* of previously-unencountered inflected forms. Nevertheless, the brief discussion offered in §7.2.3 offers insight into the direction that future research on the subject might take.

8.

CONCLUSIONS

8.0 CONCLUSIONS

This study has aimed to offer insight into why languages exhibit different diachronic developments in respect of their case systems. It has done so from an evolutionary perspective - one that centralizes the role of the speaker as the source and propagation of change in language - and has stressed the importance of taking into consideration a complex network of factors in tracing and offering explanations for the developments found in the Italic and Finnic case systems.

In chapters two and three it was demonstrated that internal factors (e.g., phonological, analogical, and syntactic considerations) and external factors (e.g., the interference from contact with other languages) do not adequately account for the changes found in the case systems of the Italic and Finnic varieties under consideration. Direct comparison of the inflexional structures found in these languages, however, reveals striking differences about the types of informative morphological patterns that can be exhibited by a case system, as was explored in chapter four.

Drawing on these conclusions, it was suggested in chapter five that there is a qualitative difference in the way that the language-user exploits instances of direct mapping of form~function mapping (substance entities) and instances in which this mapping is 'interrupted' in some way (schematic patterns). In particular, where a language is characterized by sub-word substance entities, a language-user is provided with multiple means of producing a novel form; in those languages without sub-word substance entities, frequently non-minimally-entropic schematic patterns alone must be used, which can lead to the deduction of 'incorrect' forms.

Chapter six drew together the information and arguments offered in the first five chapters of the study and applied them to the diachronic development of the Italic and Finnic case systems. It was concluded that understanding of substance entities and schematic patterns and the ways in which they are utilized by language-users in the production of novel inflected forms offers us insight into why languages lose, maintain, or propagate inflexional case-forms.

It was further argued that the role played by substance entities and schematic patterns in language production sheds light on the cross-linguistic implicational relationship between morphological structure and the number of inflexional case-forms maintained in the language. Chapter seven tested the hypothesis that agglutinative structure is a pre-requisite for a 'large' case system - that is, a case system with more than eight case values distinguished morphologically - with reference to data drawn from the *World Atlas of Linguistic Structures* online. The vast majority of languages considered in this chapter accord with the observations drawn from consideration of the Italic and Finnic case systems, although there is scope for further research to handle apparent counter-examples.

This study has revealed, and offered an explanation for, a robust and significant cross-linguistic generalization about case systems in terms of both their synchronic states and their diachronic developments. Although there is work still to be done, the observations drawn from a cross-linguistic sample of languages contribute to a wider understanding of how morphological case systems can change through time.

REFERENCES

ABONDOLO, D. (ed.). 2006a. *The Uralic Languages*. New York: Routledge.

----- 2006b. 'Introduction'. in Abondolo (ed.).1-42.

----- 2006c. 'Finnish'. in Abondolo (ed.). 149-183.

ACKERMAN, F., J.P. BLEVINS, and R. MALOUF 2009. 'Parts and wholes: Implicative patterns in inflectional paradigms'. In J.P. Blevins and J. Blevins (eds.). *Analogy in Grammar: Form and Acquisition*. Oxford: Oxford University Press. 54-81.

ACKERMAN, F. and R. MALOUF 2013. 'Morphological Organization: the Low Conditional Entropy Conjecture'. In *Language* 89:3. 429-464.

ACKERMAN, F. and R. MALOUF 2015. 'The No-Blur Principle effects as an emergent property of language systems'. In A.E. Jurgensen, H. Sande, S. Lamoureux, K. Baclawski, and A. Zerbe (eds.). *Proceedings of the 41st Annual Meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society. 1-14.

ACKERMAN, F., G.T. STUMP and G. WEBELHUTH (to appear). 'Lexicalism, periphrasis and implicative morphology'. In B. Borsley and K. Börjars (eds.). *Non-Transformational Theories of Grammar*.

ACQUAVIVA, P. 2008. *Lexical Plurals.* Oxford: Oxford University Press.

ADAMS, J.N. 2003. *Bilingualism and the Latin language.* Cambridge: Cambridge University Press.

----- 2007. *The Regional Diversification of Latin 200BC - AD600*. Cambridge: Cambridge University Press.

----- 2013. *Social Variation and the Latin Language.* Cambridge: Cambridge University Press.

AHLQVIST, A. 1863. *Suomalainen runousoppi kielelliseltä kannalta*. PhD Thesis, Helsinki.

AIKIO, A. and J. YLIKOSKI 2007. 'Suopmelaš *l*-kásusiid álgovuđđu sáme- ja eará fuolkegielaid čuovggas' ('the origin of the Finnic external local cases in light of Saami and other Uralic languages'). In J. Ylikoski and A. Aikio (eds.). *Sámit, sánit, sátnehámit. Riepmočála Pekka Sammallahtii miessemánu 21. beaivve 2007*. Mémoires de la Société Finno-Ougrienne 253. Helsinki: Finno-Ugrian Society. 11-71.

AIKIO, A. and J. YLIKOSKI. 2016. 'The origin of the Finnic *l*-cases'. In *Fenno-Ugrica Suecana Nova Series* 15. 59-158.

AINSWORTH, Z-J. 2019a. 'The Veps illative: the applicability of an abstractive approach to an agglutinative language'. In *Transactions of the Philological Society* 117:1. 58-78.

----- 2019b. 'Declension classes in Livonian - a 'language-user' abstractive approach'. In *Linguistica Uralica* 55:1. 1-21.

ALVRE, P. 1974. *Terminatiivseist konstrukstioonidest terminatiivini: Terminatiivi tulek*. Tallinn: Keel ja Kirjandus.

ANDERSON, S.R. 1992. *A-Morphous Morphology*. Cambridge: Cambridge University Press.

-----, 2015. 'The Morpheme: Its Nature and Use'. In M. Baerman (ed.). *The Oxford Handbook of Inflection*. Oxford: Oxford University Press. 11-33.

APPEL, C. 1930. *Provenzalische Chrestomathie*. Leipzig: Reisland.

ARGUS, R. 2007. *Acquisition of Morphology in Estonian.* Tallinn: Tallinn University Press.

----- 2009. ‘The early development of case and number in Estonian’. In U. Stephany and M.D. Voeikova (eds.). *Development of Nominal Inflexion in First Language Acquisition. A Cross-Linguistic Perspective*. Berlin: Mouton de Gruyter. 111-151.

----- 2012. ‘Emergence and Early Acquisition of Adjective Inflection in Estonian’. In *Journal of Baltic Studies* 43:2. 219-238.

ARONOFF, M. 1994. *Morphology by Itself.* Cambridge: The MIT Press.

AUSTERLITZ, R. 1993. ‘Agglutinative, demography, Finnish, Turkic’. In M. Sz. Bakró-Nagy and E. Szíj (eds.). *Hajdú Péter 70 éves* [Festschrift for Péter Hajdú]. Linguistica, series A, Studiae et Dissertationes 15. Budapest: MTA Nyelvtudomány intézet.

BAAYEN, R.J. 1992. ‘Quantitative aspects of morphological productivity’. In G.E. Booij and J. van Marle (eds.). *Yearbook of Morphology 1991*. Dordrecht: Kluwer Academic Publishers. 181-208.

----- 2003. ‘Probabilistic approaches to morphology’. In Bod, R., J. Hay, and S. Jannedy (eds.). *Probabilistic Linguistics*. Cambridge: The MIT Press. 229-287.

BAAYEN, R.M., J.M. MCQUEEN, T. DIJKSTRA and R. SCHREUDER 2003. ‘Frequency effects in regular inflectional morphology: Revisiting Dutch plurals’. In R.H. Baayen and R. Schreuder (eds.). *Morphological Structure in Language Processing.* Berlin: Mouton de Gruyter. 355-370.

BATEMAN, N. and M. POLINSKY 2010. ‘Romanian as a Two-Gender Language’. In D. Gerdts, J. Moore, and M. Polinsky (eds.). *Hypothesis A/Hypothesis B.* Cambridge: Cambridge University Press. 41-77.

BECKNER, C., R. BLYTHE, J. BYBEE, M.H. CHRISTIANSEN, W. CROFT, N.C. ELLIS, J. HOLLAND, J. KE, D. LARSEN-FREEMAN, and T. SCHOENEMANN 2009. 'Language Is a Complex Adaptive System: Postition Paper'. In *Language as a Complex Adaptive System.* N.C. Ellis and D. Larsen-Freeman (eds.). Chichester: Wiley-Blackwell. 1-26.

BÉDIER, J. 1927. *La Chanson de Roland, commentée par Joseph Bédier*. Paris: Piazza.

BERTINETTO, P.M. and M. SQUARTINI 2016. 'Tense and aspect'. In Ledgeway and Maiden (eds.). 939-953.

BERTACCA, A. 2009. *Natural Morphology and the Loss of Nominal Inflections in English*. Pisa: Pisa University Press.

BERTRAM, R., LAINE, M., BAAYEN, R.H., SCHREUDER, R. and HYÖNÄ, J. 2000 'Affixal homonymy triggers full-form storage even with inflected words, even in a morphologically rich language'. In *Cognition* 74:2. B13-B25.

BLAKE, B.J. 1994. *Case*. Cambridge: Cambridge University Press.

BLEVINS, J.P. 2004. 'Inflection classes and economy' in *Explorations in Nominal Inflection*. L. Gunkel, G. Müller, and G. Zifonun, (eds.). Berlin: Mouton de Gruyter. 41-85.

----- 2006. 'Word-Based Morphology'. In *Journal of Linguistics* 42:3. 531-573.

----- 2008. 'Declension classes in Estonian'. In *Linguistica Uralica* 44:4. 241-267.

----- 2009. 'Modelling analogy as probabilistic grammar'. In J. P. Blevins and J. Blevins (eds.). *Analogy in Grammar: Form and Acquisition*. Oxford: Oxford University Press. 185-213.

----- 2016. *Word and Paradigm Morphology*. Oxford: Oxford University Press.

BLEVINS, J.P., P. MILIN, and M. RAMSCAR 2017. 'The Zipfian Paradigm Cell Filling Problem'. In F. Kiefer, J.P. Blevins, and H. Bartos (eds.). *Perspectives on Morphological Structure: Data and Analyses*. Leiden: Brill. 139-158.

BLOKLAND, R. and C. HASSELBLATT 2003. 'The endangered Uralic languages'. In M. Janse and T. Sijman (eds.). *Language Death and Language Maintenance: Theoretical, practical and descriptive approaches*. The Netherlands: John Benjamins Publishing Company. 107-141.

BLOOMFIELD, L. 1933. *Language*. Chicago: University of Chicago Press.

BOD, R. (ed.) 2003. *Probabilistic Linguistics.* London: MIT Press.

BONAMI, O. and S. BENIAMINE. 2016. 'Joint predictiveness in inflectional paradigms'. In *Word Structure* 9:2. 156-182.

BONAMI, O. and J. STRNADOVÁ 2018. 'Paradigm structure and predictability in derivational morphology'. In *Morphology* 28:2. 167-197.

BOOIJ, G. 1994. 'Against split morphology'. In *Yearbook of Morphology 1993.* Dordrecht: Kluwer. 27-49.

----- 1996. 'Inherent versus contextual inflection and the split morphology hypothesis'. In G. Booij and J. van Marle (eds.). *Yearbook of Morphology 1995.* Dordrecht: Kluwer. 1-16.

BRUNEL, C. 1943. *Jaufré: roman arthurien du XIIIe siècle en vers pronvençaux. Tome I.* Paris: Société des anciens textes français.

BUTT, M. 2006. *Theories of Case.* Cambridge: Cambridge University Press.

BUTTERWORTH, B. 1983. 'Lexical representation'. In B. Butterworth (ed.). *Development, writing and other language processes, Vol. 2.* London: Academic Press.

BYBEE, J.L. 1985. *Morphology: a study of the relation between meaning and form.* Amsterdam: John Benjamins Publishing Company.

CAPRA, F. 2015. 'The Systems View of Life: A Unifying Conception of Mind, Matter, and Life'. In *Cosmos and History: The Journal of Natural and Social Philosophy* 11:2. 242-249.

CAPRA, F. and P.L. LUISI 2014. *The Systems View of Life: A Unifying Vision*. Cambridge: Cambridge University Press.

CARSTAIRS-MCCARTHY, A. 1994. 'Inflection Classes, Gender, and the Principle of Contrast'. In *Language* 70:4. 737-788.

----- 2010. *The Evolution of Morphology*. Oxford: Oxford University Press.

CHIVU, G., M. COSTINESCU, C. FRÂNCU, I. GHEȚIE, A.R. MORARU and M. TEODORESCU (eds.) 1997. *Istoria limbii române literare: epoca veche, 1532-1780.* Bucharest: Editura Academiei Române.

CLACKSON, J. and G. HORROCKS 2007. *The Blackwell History of the Latin Language.* Oxford: Blackwell Publishing.

CLARK, E. 1987. 'The Principle of Contrast: A constraint on language acquisition'. In B. MacWhinney (ed.). *Mechanisms of language acquisition*. Erlbaum: Hilsdale. 1-33.

CLARK, H.H. 1996. *Using language*. Cambridge: Cambridge University Press.

COLLINDER, B. 1960. *Comparative grammar of the Uralic languages*. Stockholm: Almqvist and Wiksell.

----- 1965. *An Introduction to the Uralic Languages.* Berkeley: University of California Press.

CORBETT, G.G. 2007. ‘Canonical typology, suppletion, and possible words’. In *Language* 83:1. 8-42.

----- 2009. ‘Canonical Inflectional Classes’. In F. Montermini, G. Boyé, and J. Tseng (eds.). *Selected Proceedings of the 6th Décembrettes*. Somerville: Cascadilla Proceedings Project. 1-11.

CORBIN, D. 1987. ‘Morphologie dérivationnelle et structuration du lexique’. Volumes I-II. *Linguistische Arbeiten* 193. Tübingen: Max Niemeye Verlag.

CORNILESCU, A. and A. NICOLAE 2011. ‘On the Syntax of Romanian Definite Phrases: Changes in the Patterns of Definiteness Marking’. In P. Sleeman and H. Perridon (eds.). *The Noun Phrase in Romance and Germanic. Structure, Variation and Change.* Amsterdam: John Benjamins. 193-221.

COŞERIU, E. 1987. ‘Le latin vulgaire et le type linguistique roman: à propos de la thèse de Humboldt: ‘Es sanken Formen, nicht aber die Form’.’ In J. Herman (ed.). *Latin vulgaire - latin tardif: actes du 1er Colloque international sur le latin vulgaire et tardif (Pecs, 2-5 Sept. 1985).* Tübingen: Niemeyer. 53-64.

CRIPPEN, J.A. 2010. *Lingít Yoo X̲’atángi. A Grammar of the Tlingit Language*. Honolulu: The Crippen Press.

CROFT, W. 2000. *Explaining Language Change*. Essex: Pearson Education Limited.

CRUSCHINA, S. 2016. ‘Information and discourse structure’. In Ledgeway and Maiden (eds.). 596-608.

DANILIUC, L. and R. DANILIUC. 2000. *Descriptive Romanian Grammar: an outline.* Muenchen: Lincom Europa.

DENCH, A.C. 1987. *Martuthunira: A Language of the Pilbara Region of Western Australia.* Graduate thesis, The Australian National University.

DOTTIN, G. 1918. *La Langue Gauloise: Grammaire, Textes et Glossaire*. Paris: Librairie C. Klincksieck.

DRAGER, K.K. 2011. 'Sociophonetic variation and the lemma'. In *Journal of Phonetics* 39:4. 694-707.

DRAGOMIRESCU, A. and A. NICOLAE. 2016. 'Case'. In Ledgeway and Maiden (eds.). 911-923.

DRESSLER, W.U. 2003. 'Morphological Typology and First Language Acquisition: Some Mutual Challenges'. In G. Booij, E. Guevara, A. Ralli, S. Sgroi, and S. Scalise (eds.). *Morphology and Linguistic Typology, On-line Proceedings of the Fourth Mediterranean Morphology Meeting (MMM4) Catania 21-23 September 2003*. University of Bologna 2005. 7-20.

DUNN, M.J. 1999. *A Grammar of Chukchi*. Graduate thesis: Australian National University.

EHALA, M. 1994. 'Russian influence and the change in progress in the Estonian adpositional system'. In *Linguistica Uralica* 3. 177-193.

-----, 2009. 'Linguistic strategies and markedness in Estonian morphology'. In *Language Typology and Universals* 62:1-2. 29-48.

EISENBEISS, S., B. NARASHIMHAN, and M. VOEIKOVA. 2011. 'The acquisition of case'. In Malchukov and Spencer (eds.). 369-383.

ERNŠTREITS, V. 2007. 'Livonian Orthography'. In *Linguistica Uralica* 43:1. 11-22.

ERNŠTREITS, V. and G. KĻAVA. 2014. ‘Grammatical changes caused by contact between Livonian and Latvian’. In *Eesti ja some-ugri keeleteaduse ajakiri* 5:1. 77-90.

EVANS, N. and S.C. LEVINSON 2009. ‘The myth of language universals: Language diversity and its importance for cognitive science’. In *Behavioral and Brain Sciences* 32:5. 429-492.

FELDMAN, L.B. 1991. ‘The Contribution of Morphology to Word Recognition’. In *Psychological Research* 53:1. 33-41.

FOULET, L. 1970. *Petite Syntaxe de l’ancien français*. Paris: H. Champion.

GAENG, P.A. 1977. *A Study of Nominal Inflections in Latin Inscriptions: A Morpho-Syntactic Analysis*. Chapel Hill: University of North Carolina Studies in the Romance Languages and Literatures 182.

GAHL, S. 2008. ‘“Thyme” and “Time” are not homophones. The effect of lemma frequency on word durations in spontaneous speech’. In *Language* 84:3. 474-496.

GARDANI, F. 2018. ‘On morphological borrowing’. In *Language and Linguistics Compass* 12:10. 1-17.

GILES, H. and P.M. SMITH 1979. ‘Accommodation theory: optimal levels of convergence’. In H. Giles and R. N. St Clair (eds.). *Language and social psychology.* Oxford: Basil Blackwell.

GIURGEA, I. 2013. ‘Nominal inflection’. In C. Dobrovie-Sorin and I. Giurgea (eds.). *A Reference Grammar of Romanian.* Amsterdam: John Benjamins Publishing Company. 827-856.

GÖNCZÖL, R. 2008. *Romanian: An Essential Grammar*. London: Routledge.

GRANDGENT, C.H. 1909. *An Outline of the Phonology and Morphology of Old Provençal*. Boston: D.C. Heath & Co. Publishers.

GREENBERG, J.H. 1963. 'Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements'. In J.H. Greenberg (ed.). *Universals of language: report of a conference held at Dobbs Ferry, New York, April 13-15, 1961*. Cambridge, Mass.: MIT Press. 73-113.

GRÜNTHAL, R. 2003. *Finnic Adpositions and Cases in Change.* Helsinki: Société Finno-Ougrienne.

----- 2005. 'Miksi itämerensuomessa on prepositioita?'. In *Virittäjä* 109:1. 28-51.

----- 2015a. *Vepsän kielioppi*. Helsinki: Suomalais-Ugrilainen Suera.

----- 2015b. 'Livonian at the crossroads of language contacts'. In S. Junttila (ed.). *Contacts between the Baltic and Finnic* languages. Helsinki: University of Helsinki. 97-150.

----- 2017. 'The essive in Veps'. In C. de Groot (ed.). *Uralic Essive and the Expression of Impermanent State*. Amsterdam: John Benjamins Publishing Company. 131-159.

GUȚU ROMALO, V. 1968. *Morfologie structurală a limbii române (substantiv, adjectiv, verb).* Bucharest: Editura Academiei Romane.

HAAVIO, M. 1964. 'Tuohikirje n:o 292: Vanha suomalaisen muinaisuskonnon lähde'. In *Virittäjä* 68:1. 1-17.

HAGÈGE, C. 2010. *Adpositions*. Oxford: Oxford University Press.

HAKANEN, J. 2006. *Äyräitä vai äyräksiä? Suomen kielen s-loppuisten nominien taivutus S2-oppijoilla ja natiiveilla.* Graduate thesis, University of Tampere.

HÄKKINEN, K. 2004. *Nykysuomen etymologinen sanakirja*. Helsinki: Werner Söderström OY.

HAKULINEN, L. 1979. *Suomen kielen rakenne ja kehitys.* Helsinki: Otava.

HÄMÄLÄINEN, M. and F. ANDREJEV 1934. *Vepsikijan kelen grammatik.* Leningrad: Kirja.

HALLE, M. and A. MARANTZ, 1993. 'Distributed Morphology and the pieces of inflection'. In K. Hale and S. J. Keyser (eds.). *The View from Building 20: Linguistic Essays in Honor of Sylvain Bromberger*. Cambridge, MA: MIT Press. 111-176.

HASPELMATH, M. 1993. *A Grammar of Lezgian*. Berlin: Mouton de Gruyter.

HEINE, B. and H. NARROG, 2011. 'The grammaticalization of adpositions and case marking'. In C. König (ed.) *The Oxford Handbook of Grammaticalization*. Oxford: Oxford University Press. 511-521.

HELIMSKI, E. 1983. *The Language of the First Selkup Books*. Studia Uralo-Altaica. Szeged: József Attila University.

-----, 2006. 'Selkup'. In Abondolo (ed.). 548-579.

HERMAN, J. 2000. *Vulgar Latin*. Pennsylvania: The Pennsylvania State University Press. (Translator: R. Wright).

HILL, V. and G. ALBOIU, 2016. *Verb Movement and Clause Structure in Old Romanian*. Oxford: Oxford University Press.

HOCKETT, C.F. 1954. 'Two models of grammatical description'. In *Word* 10:2-3. 210-231.

IGARTUA, L. 2006. '*Genus alternans* in Indo-European'. In *Indogermanische Forschungen* 111:1. 56-70.

IGGESEN, O.A. 2011. ‘Asymmetry in Case Marking: Nominal vs. Pronominal Systems’. In Malchukov and Spencer (eds.). 246-257.

INABA, N. 2015. *Suomen datiivigenetiivin juuret vertailevan menetelmän valossa.* Helsinki: Suomalais-Ugrilainen Seura.

ISK = A. Hakulinen, M. Vilkuna, R. Korhonen, V. Koivisto, T.R. Heinonen, and I. Alho (eds.) 2004. *Iso Suomen Kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.

ITKONEN, T. 1966. *Kieli ja sen tutkimus*. Helsinki: W. Söderström.

----- 1977. ‘Huomioita lapsen ääneistön kehityksestä’. In *Virittäjä* 81:3. 279-308.

JENSEN, F. 1976. *The Old Provençal Noun and Adjective Declension.* Denmark: Odense University Press.

de JONG, N.J., R. SCHREUDER and R.H. BAAYEN 2000. ‘The morphological family size effect and morphology’. In *Language and cognitive processes*. 15:4-5. 329-365.

KABAK, B. 2006. ‘An obstacle to the morphologization of postpositions’. In *Studies in Language* 30:1. 33-68.

KAHNEMAN, D. 2011. *Thinking, Fast and Slow.* New York: Farrar, Straus and Giroux.

KAHR, J.C. 1976. ‘The Renewal of Case Morphology: Sources and Constraints’. In *Working Papers on Language Universals* (Stanford University) 20. 107-151.

KANERVA, J. 1987. ‘Morphological integrity and syntax: The evidence from Finnish possessive suffixes’. In *Language* 63:3. 498-521.

KARLSSON, F. 1982. *Suomen kielen äänne- ja muotorakenne*. Porvoo: Söderström.

----- 2006. ‘Finnish as an agglutinating language’. In K. Brown (ed.). *Elsevier Encyclopedia of Language & Linguistics*. Oxford: Elsevier. 476-480.

----- 2008. *Finnish: An Essential Grammar.* London: Routledge.

KEMPS, R.J.J.K, M. ERNESTUS, R. SCHREUDER, and R.H. BAAYEN 2005. ‘Prosodic cues for morphological complexity: The case of Dutch plural nouns’. In *Memory & Cognition* 33:3. 430-446.

KETTUNEN, L. 1929. *Eestin kielen äännehistoria. Toinen, uusittu painos.* Helsinki: Suomalaisen Kirjallisuuden Seuran Toimituksia 156.

-----, 1938. *Livisches Wörterbuch mit grammatischer Einleitung.* Helsinki: Suomalais-Ugrilainen Seura.

----- 1943. *Vepsän murteiden lauseopillinen tutkimus*. Helsinki: Suomalais-Ugrilainen Seura.

----- 1947. *Hauptzüge der livischen Laut- und Formengeschichte.* Helsinki: Suomalais-Ugrilainen Seura.

KIHM, A. 2012. ‘Old French and Romanian Declensions from a *Word and Paradigm* Perspective and the Notion of “Default Syncretism”’. In *Revue Roumaine de Linguistique* 57:1. 3-34.

----- 2017. ‘Old French declension: a Word and Paradigm approach and the role of syncretisms and defaults in its rise and fall’. In N. Gisbourne and A. Hippisley (eds.). *Defaults in Morphological Theory*. Oxford: Oxford University Press. 40-72.

KITTILÄ, S. and J. YLIKOSKI 2011. 'Remarks on the coding of Goal, Recipient and Vicinal Goal in European Uralic'. In Kittilä, S., K. Västi, and J. Ylikoski (eds.). *Case, Animacy and Semantic Roles*. Amsterdam: John Benjamins Publishing Company. 29-65.

KITTILÄ, S., K. VÄSTI, and J. YLIKOSKI 2011. 'Introduction to case, animacy and semantic roles'. In Kittilä, S., K. Västi, and J. Ylikoski (eds.). *Case, Animacy and Semantic Roles*. Amsterdam: John Benjamins Publishing Company. 1-26.

KÖNIG, C. 2011. 'The grammaticalization of adpositions and case marking'. In B. Heine and H. Narrog (eds.). *The Oxford Handbook of Grammaticalization*. Oxford: Oxford University Press. 511-521.

KORHONEN, M. 1996a. 'Typological drift in the Finno-Ugrian languages with special reference to the case system'. In *Typological and Historical Studies in Languages by Mikko Korhonen: a memorial volume published on the 60th anniversary of his birth*. T. Salminen (ed.). Helsinki: Suomalais-Ugrilainen Seura. 195-206.

----- 1996b. 'Remarks on the structure and history of the Uralic case system'. In *Typological and Historical Studies in Languages by Mikko Korhonen: a memorial volume published on the 60th anniversary of his birth*. T. Salminen (ed.). Helsinki: Suomalais-Ugrilainen Seura. 219-234.

KUHL, P.A. 2000. 'A new view of language acquisition'. In *Proceedings of the National Academy of Sciences of the United States of America* 92:22. 11850-11857

KULIKOV, L. 2006. 'Case systems in a diachronic perspective: A typological sketch'. In L. Kulikov, A. Malchukov, and P. de Swart (eds.). *Case, Valency and Transitivity*. Amsterdam: John Benjamins Publishing Company. 23-47.

-----, 2011. 'Evolution of Case Systems'. In Malchukov and Spencer (eds.). 439-457.

LAAKSO, J. 2001. 'The Finnic languages'. In O. Dahl and M. Koptjevskaja-Tamm (eds.). *Circum-Baltic Languages. Volume 1: Past and Present.* Amsterdam: John Benjamins Publishing Company. 179-212.

LAALO, K. 2009. 'Acquisition of Case and Plural in Finnish'. In U. Stephany and M. Voeikova (eds.). *Development of Nominal Inflection in first Language Acquisition: a Cross-Linguistic Perspective.* Berlin: Mouton de Gruyter. 49-90.

LAINE, M., J. NIEMI, P. KOIVUSELKA-SALLINEN, and J. HYÖNÄ 1995. 'Morphological processing of polymorphemic nouns in a highly inflecting language'. In *Cognitive Neuropsychology* 12:5. 457-502.

LARSSON, L. 2001. 'Baltic influence on Finnic languages'. In O. Dahl and M. Koptjevskaja-Tamm (eds.). *Circum-Baltic Languages. Volume 1: Past and Present.* Amsterdam: John Benjamins Publishing Company. 238-253.

LASS, R. 1990. 'How to do things with junk: exaptation in language evolution'. In *Journal of Linguistics* 26:1. 79-102.

----- 1997. *Historical linguistics and language change*. Cambridge: Cambridge University Press.

LAUSBERG, H., 1973. *Lingűística Románica Tomo II: Morfología*. Madrid: Gredos.

LEDGEWAY, A. 2012. *From Latin to Romance: Morphosyntactic Typology and Change.* Oxford: Oxford University Press.

----- 2017. 'The Romanian definite article in a comparative Romance perspective'. In A. Dragomirescu, A. Nicolae, C. Stan, and R. Zafiu (eds.). *Sintaxa ca mod de a fi. Omagiu Gabrielei Pană Dindelegan, la aniversare*. Bucharest: Editura Universităţii din Bucureşti.

LEDGEWAY, A. and M. MAIDEN (eds.). 2016. *The Oxford Guide to the Romance Languages.* Oxford: Oxford University Press.

LIEBER, R. 1980. *On the organization of the lexicon*. Doctoral dissertation, MIT. [Reproduced by Indiana University Linguistics Club, 1981.]

-----, 1988. 'Configurational and nonconfigurational morphology'. In M. Everaert, A. Evers, R. Huybregts, and M. Trommelen (eds.). *Morphology and modularity*. Dordrecht: Foris. 187-215.

-----, 1992. *Deconstructing morphology: Word formation in syntactic theory.* Chicago: University of Chicago Press.

LÖNNROT, E. 1841. *Elias Lönnrotin matkat II: 1841-1844*. Helsinki: Suomalaisen Kirjallisuuden Seura.

LOPORCARO, M. 2017. *Gender from Latin to Romance: History, Geography and Typology*. Oxford: Oxford University Press.

LOT, F. 1931. 'A quelle époque a-t-on cessé de parler latin?'. In *Archivum Latinitatis Medii Aevi 6*. 97-159.

MAAMIES, S. 2015. '*Kompromisseja* vai *kompromissejä* - vierassanojen taivutuspäätteen vokaali'. In *Kielikello* 1. 13-17.

MAIDEN, M. 1996. 'On the Romance inflectional endings *-i* and *-e*'. In *Romance Philology* 50. 147-182.

-----, 2007. 'Lexical nonsense and morphological sense'. In T. Eythósson (ed.). *Grammatical Change and Linguistic Theory: The Rosendal papers*. Amsterdam: John Benjamins Publishing Company. 307-328.

----- 2011a. ‘Morphological Persistence’. In Maiden, Smith, and Ledgeway (eds.). 155-215.

----- 2011b. ‘Morphophonological Innovation’. In Maiden, Smith, and Ledgeway (eds.). 216-267.

----- 2013. ‘The Latin ‘Third Stem’ and its Romance Descendants’. In *Diachronica* 30:4. 492-530.

----- 2015. ‘The plural type *cărnuri* and the morphological structure of the Romanian feminine noun in diachrony’. In G. Pană Dindelegan, R. Zafiu, A, Dragomirescu, I. Nicula, A. Nicolae, and L. Escher (eds.). *Diachronic Variation in Romanian*. Newcastle upon Tyne: Cambridge Scholars. 35-54.

----- 2016a. ‘Romanian, Istro-Romanian, Megleno-Romanian, and Aromanian’. In Ledgeway and Maiden (eds.). 91-125.

----- 2016b. ‘Inflectional morphology’. In Ledgeway and Maiden (eds.). 497-512.

----- 2016c. ‘Number’. In Ledgeway and Maiden (eds.). 697-707.

----- 2016d. ‘The Romanian alternating gender in diachrony and synchrony’. In *Folia Linguistica Historica* 37:1. 111-144.

----- 2016e. ‘Some lessons from history: morphomes in diachrony’. In A. Luís and R. Bermúdez-Otero (eds.). *The Morphome* Debate. Oxford: Oxford University Press. 33-63.

----- 2016f. ‘Ambiguity in Romanian Word-Structure. The Structure of Plurals in ...*uri*’. In *Revue roumaine de linguistique* 61:1. 3-12.

----- 2018. *The Romance Verb: Morphomic Structure and Diachrony*. Oxford: University of Oxford Press.

----- 2019. ‘A note on the inflexional morphology of the Vegliote Dalmatian masculine plural’. In *Philologica Jassyensia* 29. 81-88.

MAIDEN, M., J.C. SMITH, and A. LEDGEWAY (eds.) 2011. *The Cambridge History of the Romance Languages.* Cambridge: Cambridge University Press.

MAKAROV, G.N. 1963. *Obrazcy karel’skoj reči. Kalininskie govory.* Leningrad: Moskva.

MALCHUKOV, A. and A. SPENCER (eds.) 2011. *The Oxford Handbook of Case*. Oxford: Oxford University Press.

MANU MAGDA, M. 2013. ‘Vocative phrases and address’. In Pană Dindelegan (ed.). 590-598.

MARTIN, M. 1995. *The Map and the Rope: Finnish nominal inflection as a learning target*. Jyväskylä: University of Jyväskylä.

MATHIASSEN, T. 1997. *A Short Grammar of Latvian.* Columbus, Ohio: Slavica Publishers.

MATTHEWS, P.H. 1972. *Inflectional morphology: a theoretical study based on aspects of Latin verb conjugation.* Cambridge: Cambridge University Press.

MEILLET, A. 1903. *Introduction à l’étude comparative des language indo-européennes.* Paris: Hachette.

MEYER-LÜBKE, W. 1895. *Grammaire des langues romanes: morphologie.* Paris: Welter.

MIKLOSICH, F. 1882. *Beiträge zur Lautlehre der rumunischen Dialekte. Vocalismus III. Consonantismus I.* Vienna: Kaiseliche Akademie der Wissenschaften.

MOIGNET, G. 1976. *Grammaire de l’ancien français*. Paris: Klincksieck.

NEDELCU, I. 2013a. ‘Three Genders: Masculine, Feminine, and Neuter’. In Pană Dindelegan (ed.). 255-258.

----- 2013b. ‘Inflection classes’. In Pană Dindelegan (ed.). 273-277.

----- 2013c. ‘Nouns with irregular inflection’. In Pană Dindelegan (ed.). 277-278.

----- 2013d. ‘Prepositions and Prepositional Phrases’. In Pană Dindelegan (ed.). 451-465.

NESSER, A. 1981. *Finska barns språkutveckling: fonologi och morfologi*. University of Uppsala: Finsk-ugriska institutionen.

NICHOLS, J. 1986. ‘Head-Marking and Dependent-Marking Grammar’. In *Language* 62:1. 56-119.

----- 2011. *Ingush Grammar*. London: University of California Press.

NIEMI, J. and S. NIEMI 1987. ‘Acquisition of inflectional marking: A case study of Finnish’. In *Nordic Journal of Linguistics* 10:1. 59-89.

NORDLINGER, R. 1998. *A Grammar of Wambaya, Northern Australia*. Canberra: Pacific Linguistics.

NOYER, R.R. 1992. *Features, positions and affixes in autonomous morphological structure*. Doctoral dissertation, MIT.

NYROP, K. 1924. *Grammaire historique de la langue française Vol. II.* Paris: Alphonse Picard & fils.

----- 1935. *Grammaire historique de la langue française Vol. I*. Paris: Alphonse Picard & fils.

OINAS, F. 1961. *The Development of Some Postpositional Cases in Balto-Finnic Languages*. Helsinki: Suomalais-Ugrilainen Seura.

OJAJÄRVI, A. 1950. *Sijojen merkitustehtävistä Itä-Karjalan Maaselän murteissa.* Helsinki: Suomalais-Ugrilainen Seura.

PADEN, W.D. 1998. *An Introduction to Old Occitan*. New York: Modern Language Association of America.

PAJUSALU, K. 2012. 'Phonological Innovations of the Southern Finnic Languages'. In G. Riho and P. Kallio (eds.). *A Linguistic Map of Prehistorical Northern Europe*. Helsinki: Suomalais-Ugrilaisen Seura. 201-224.

----- 2014. 'The Relationship between Salaca Livonian and Courland Livonian Dialects'. In *Eesti ja soome-ugri keeleteaduse ajakiri* 5:1. 149-171.

PALMER, L.R., 1954. *The Latin Language*. London: Faber and Faber Ltd.

PANĂ DINDELEGAN, G. (ed.). 2013a. *The Grammar of Romanian.* Oxford: Oxford University Press.

----- 2013b. 'Introduction'. In Pană Dindelegan (ed.). 1-17.

----- 2013c. 'The Structure of Root Clauses'. In Pană Dindelegan (ed.). 100-125.

PĂTRUȚ, I. 1956. 'Există gen 'neutru' în limba româna'. In *Cercetări de lingvistică* 1:29. 29-40.

PEIRCE, C.S. 1998 [1894]. 'What Is a Sign?' In *The Essential Peirce: selected philosophical writings*. Houser, Nathan (ed.). Bloomington: Indiana University Press. 4-10.

PENNEY, J. 2011. 'Archaic and Old Latin'. In J. Clackson (ed.). *A Companion to the Latin Language*. Oxford: Wiley-Blackwell. 220-235.

PENNY, R. 2002. *A History of the Spanish Language*. Cambridge: Cambridge University Press.

PHILIPPIDE, A. 2011. *Istoria limbii române*. Iaşy: Polirom.

PINKSTER, H. 1990. 'The development of cases and adpositions in Latin'. In H. Pinkster and I. Genee (eds.). *Unity in Diversity: Papers Presented to Simon C. Dik on his 50th Birthday*. Holland: Foris Publications. 195-209.

----- 2015. *The Oxford Latin Syntax. Volume 1: The Simple Clause.* Oxford: Oxford University Press.

PLAG, I., J. HOMANN, and G. KUNTER. 2017. 'Homophony and morphology: The acoustics of word-final S in English'. In *Journal of Linguistics* 53:1. 181-216.

PLANK, F. 1986. 'Paradigm size, morphological typology, and universal economy'. In *Folia Linguistica* 20:1-2. 29-48.

POPE, M.K. 1952. *From Latin to Modern French with especial consideration of Anglo-Norman: phonology and morphology.* Manchester: Manchester University Press.

POSNER, R. 1996. *The Romance Languages*. Cambridge: Cambridge University Press.

del PRADO MARTÍN F.M., R. BERTRAM, T. HÄIKIÖ, R. SCHREUDER, and R.H. BAAYEN, 2004. 'Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew'. In *Journal of Experimental Psychology, Language Memory and Cognition*. 30:6. 1271-1279.

PRICE, G. 1971. *The French language: present and past*. London: Edward Arnold.

PUURA, U., H. KARJALAINEN, N. ZAJCEVA, and R. GRÜNTHAL 2013. *The Veps Language in Russia: ELDIA Case-Specific Report*. Studies in European Language Diversity 25. Mainz: Research consortium ELDIA.

QUER, J. 2016. 'Mood'. In Ledgeway and Maiden (eds.). 954-966.

RÄISÄNEN, A. 1975. 'Havaintoja lastenkielestä'. In *Virittäjä* 79:3. 251-266.

RAMSCAR, M. 2013. 'Suffixing, prefixing, and the functional order of regularities in meaningful strings'. In *Psihologija* 46:4. 377-396.

RAMSCAR, M. and M. DYE 2010. 'Learning language from the input: Why innate constraints can't explain noun compounding'. In *Cognitive Psychology* 62:1. 1-40.

RAMSCAR, M., D. YARLETT, M. DYE, K. DENNY, and K. THORPE 2010. 'The effects of feature-label-order and their implications for symbolic learning'. In *Cognitive Science* 34:6. 909-957.

RAPOLA, M. 1996. *Suomen kielen äännehistorian luennot*. Helsinki: Suomalaisen kirjallisuuden seura.

van REENEN, P. and L. SCHØSLER 1998. 'Formation and evolution of the feminine and masculine singular nouns in Old French *la maison(s)* and *li charbons*'. In J. Fisiak (ed.). *Historical Dialectology*. Berlin: Mouton de Gruyter. 505-545.

RESCORLA, R.A. 1988. 'Pavlovian conditioning: It's not what you think it is'. *American Psychologist* 43:3. 151-160.

RESCORLA, R.A. and A.R. WAGNER 1972. 'A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement'. In A. H. Black and W. F. Prokasy (eds.). *Classical Conditioning.* New York: Appleton-Century-Crofts. 64-99.

ROBINS, R.H. 1959. ‘In defence of WP’. In *Transactions of the Philological Society* 58:1. 116-144.

ROZENTĀLS, A. 2008. *Vepsu-Latviešu Vārdnīca ~ Vepsä-Latvijalaine Vajehnik*. Rīgā: Eraksti.

ROZHANSKIY, F. and E. MARKUS 2015. ‘Dialectal variation in Votic: Jõgõperä vs. Luuditsa’. In *Eesti ja soome-ugri keeleteaduse ajakiri* 6:1. 23-29.

SALVI, G. 2011. ‘Morphosyntactic persistence’. In Maiden, Smith, and Ledgeway (eds.). 318-381.

SAPIR, E. 1970. *Language: an introduction to the study of speech*. London: Hart-Davis.

SÄRKKÄ, T. 1969. *Itämerensuomalaisten kielten eksessiivi*. Helsinki: Suomalaisen Kirjallisuuden Seura.

SCHØSLER, L. 1984. *La déclinaison bicasuelle de l’ancien français*. Odense: Odense University Press.

SCHREUDER, R. and R.H. BAAYEN 1997. ‘How complex simple words can be’. In *Journal of Memory and Language*. 37:1. 118-139.

SETÄLÄ, E. N. 1891. *Yhteissuomalainen äännehistoria I-II*. Helsinki.

SIHLER, A.L. 1995. *New comparative grammar of Greek and Latin.* Oxford: Oxford University Press.

SKÅRUP, P. 1997. *Morphologie élémentaire de l’ancien occitan*. Copenhagen: Museum Tusculanum Press.

SKOUSEN, R. 1975. *Substantive Evidence in Phonology: The Evidence from Finnish and French*. Paris: Mouton.

SMITH, J.C. 2005. 'Some refunctionalizations of the nominative-accusative opposition between Latin and Gallo-Romance'. In B. Smelik, R. Hofman, C. Hamans, and D. Cram (eds.). *A Companion in Linguistics: a Festschrift for Anders Ahlqvist on the occasion of his sixtieth birthday*. Nijmegen: De Keltische Draak. 269-285.

----- 2011. 'Change and Continuity in Form-Function Relationships'. In Maiden, Smith, and Ledgeway (eds.). 268-317.

----- 2013. 'The morphome as a gradient phenomenon: evidence from Romance'. In S. Cruschina, M. Maiden, and J.C. Smith (eds.). *The Boundaries of Pure Morphology: Diachronic and Synchronic Perspectives*. Oxford: Oxford University Press. 247-261.

----- 2016. 'French and northern Gallo-Romance'. In Ledgeway and Maiden (eds.). Oxford: Oxford University Press. 292-318.

SORNICOLA, R. 2011. 'Romance Linguistics and Historical Linguistics: Reflections on Synchrony and Diachrony'. In Maiden, Smith, and Ledgeway (eds.). 1-49.

SPENCER, A. 2013. *Lexical relatedness: A paradigm-based model.* Oxford: Oxford University Press.

SPENCER, A. and A.M. ZWICKY (eds.). 2011. *The Handbook of Morphology*. Oxford: Blackwell Publishing.

STAN, C. 2013. 'The enclitic definite article. The proclitic indefinite article'. In Pană Dindelegan (ed.). 285-294.

STAN, C. and M. M. POLOGEA 2013. 'Inflectional and Derivational Morphophonological Alternations'. In Pană Dindelegan (ed.). 607-611.

STANOVAÏA, L. 1993. ‘Sur la déclinaison bicasuelle en ancien français (point de vue scriptologique)’. *Travaux de linguistique et de philologie* 23. 163-182.

STEMBERGER, J.P. and B. MACWHINNEY 1986. ‘Frequency and the lexical storage of regularly inflected forms’. In *Memory and Cognition* 14:1. 17-26.

STERIADE, D. 2016. ‘The morphome vs similarity-based syncretism’. In A. Luís and R. Bermúdez-Otero (eds.). *The Morphome Debate.* Oxford: Oxford University Press. 112-171.

STUMP, G. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

----- 2005. ‘Some criticisms of Carstairs-McCarthy’s conclusions’. In G. Booij and J. van Marle (eds.). *Yearbook of Morphology 2005*. 283-303.

STUMP, G. and R.A. FINKEL 2013. *Morphological Typology: From Word to Paradigm.* Cambridge: Cambridge University Press.

TAFT, M. 1979. ‘Recognition of affixed words and the word frequency effect’. In *Memory and Cognition* 7:4. 263-272.

TIKKA, T. 1992. *Vepsän suffiksoituneet postpositiot: Kieliopillisiin sijoihin liittyvä suffiksoituminen.* Uppsala: Studia Uralica Upsaliensia 22.

THOMASON, S.G. 2001. *Language Contact: and Introduction.* Washington, D.C.: Georgetown University Press.

THOMASON, S.G. and T. KAUFMAN 1988. *Language Contact, Creolization and Genetic Linguistics*. Los Angeles: University of Californian Press.

TRUDGILL, P. 1986. *Dialects in contact*. Oxford: Blackwell.

----- 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press.

TUISK, T. 2016. 'Main features of the Livonian sound system and pronunciation' in *Journal of Estonian and Finno-Ugric Linguistics* 7:1. 121-143.

VÄÄNÄNEN, V. 1981. *Introduction au latin vulgaire.* Paris: Klincksiek.

----- 1966. *Le latin vulgaire des inscriptions pompéiennes*. Berlin: Akademie-Verlag.

VAHTIA, E. 1938. *Karjalan kielioppi*. Suojärvi: Karjalan Kansalaisseura.

VIITSO, T-R. 2007. 'Livonian Gradation: Types and Genesis'. In *Linguistica Uralica* 43:1. 45-62.

----- 2012. 'Concerning Inflection Classes in Livonian'. In *Linguistica Uralica* 48:1. 12-26.

----- 2016. 'The Essive in Livonian'. In *Journal of Estonian and Finno-Ugric Linguistics* 7:1. 145-175.

VIITSO, T-R. and V. ERNŠTREITS 2012. *Līvõkīel-ēstikīel-leţkīel sõnārõntõz = Liivi-eesti-läti sõnaraamat = Lībiešu-igauņu-latviešu vārdīca*. Tartu: Tartu Ülikool.

VILKUNA, M. 2003. *Suomen lauseopin perusteet*. Helsinki: Edita Publishing Oy.

VINCENT, N. 2016. 'A structural comparison of Latin and Romance'. In Ledgeway and Maiden (eds.). 37-49.

VINCENT, N. and K. BÖRJARS 2010. 'Grammaticalization and models of language'. In E. Closs Traugott and G. Trousdale (eds.). *Gradience, Gradualness and Grammaticalization*. Amsterdam: John Benjamins Publishing Company. 279-299.

WAGNER, A.R. and R.A. RESCORLA 1972. ‘Inhibition in Pavlovian conditioning: Application of a theory’. In R.A. Boakes and M.S. Hallinday (eds.). *Inhibition and learning.* New York: Academic Press. 301-336.

WÄLCHLI, B. 2000. ‘Livonian in a genetic, areal and typological perspective, or is Finnish better Finnic than Livonian?’ In J. Laakso (ed.). *Facing Finnic. Some Challenges to Historical and Contact Linguistics. Symposium at the Congressus Nonus Internationalis Fenno-Ugristarum held in Tartu.* Helsinki: Castrenianumien. 210-226.

----- 2001. ‘Lexical evidence for the parallel development of the Latvian and Livonian verb particles’. O. DAHL and M. Koptjevskaja-Tamm (eds.). *Circum-Baltic Languages. Volume 2: Grammar and Typology*. Amsterdam: John Benjamins Publishing Company. 413-441.

WOLEDGE, B., J. BEARD and C.H.M. HORTON 1967-69. ‘La déclinaison des substantifs dans la Chanson de Roland recherches mécanographiques’. In *Romania* 88. 145-74; 90. 174-201.

WOODCOCK, E.C. 1959. *A New Latin Syntax.* London: Methuen.

WRIGHT, R. (ed.). 2011. ‘Romance Languages As A Source For Spoken Latin’. In J. Clackson (ed.). *A Companion to the Latin Language*. Oxford: Blackwells. 59-79.

----- 2013. ‘Periodization’. In Maiden, Smith, and Ledgeway. (eds.). 107-124.

WURZEL, W.U. 1987. ‘System-dependent morphological naturalness in inflection’. In W.U. Dressler, W. Mayerthaler, O. Panagl, and W.U. Wurzel (eds.). *Leitmotifs in Natural Morphology*. Amsterdam: John Benjamins Publishing Company. 59-96.

YLIKOSKI, J. 2016. 'The origins of the western Uralic *s*-cases revisited: historiographical, functional-typological and Samoyedic perspectives'. In *Finnisch-Ugrische Forschungen* 63:1. 6-78.

ZAJCEVA, M. 1981. *Именное словоизменение в вепсском языке: (история и функционирование форм слова)*. Karelija: Izdat.

ZAJCEVA, N. 1995. *Vepsän kelen grammatik I. Nimiden kändluz.* Petrozavodsk: Izdate'stvo Petrozavodskogo gosudarstvennogo universiteta.

ZAJKOV, P.M. 1999. *Grammatika karel'skogo jazyka: fonetika i morfologija*. Petrozavodsk: Periodika.

ZINK, G. 1989. *Morphologie du français medieval.* Paris: Presses Universitaires de France.

ZIPF, G.K. 1935. *The Psychobiology of Language: An Introduction to Dynamic Philology.* Cambridge: MIT Press.

----- 1949. *Human Behaviour and the Principle of Least Effort*. Cambridge: Addison-Wesley.