Divergence and Mismatches in Portuguese Morphology and Syntax



A thesis submitted for the degree of Doctor of Philosophy
in Comparative Philology and General Linguistics

Amanda L. R. Thomas

Trinity College, University of Oxford

Trinity Term 2021

Contents

List of Figures v
List of Tables vi
List of Abbreviations viii
Abstract xii
Acknowledgements xiii
Chapter 1 Introduction 1
1.1 Research questions 2
1.2 Novel aspects of this research 3
1.3 Brief introduction to Portuguese 7
1.4 Structure of the thesis 12
Chapter 2 Theoretical background 16
2.1 Lexical-Functional Grammar 16
2.1.1 Fundamentals of LFG 16
2.1.2 Advantages of LFG and justification for its use 22
2.2 Paradigm Function Morphology 25
2.2.1 Fundamentals of PFM 26
2.2.2 Advantages of PFM and justification for its use 30
2.3 Combining LFG and PFM 32
2.3.1 Lexemic entries 34
2.3.2 Lexical entries 36
2.3.3 M-entries and the morphological realisation relation $\mathscr{R}$ 39
2.3.4 The description function $\mathcal{D}$ 41
2.3.5 The relationship between f-features and m-features 43
2.3.6 Categories in LFG and PFM 45
2.4 Chapter summary 47
Chapter 3 Applying LFG and PFM to Portuguese 49
3.1 Analysis of Portuguese syntax and morphology 49
3.1.1 Basics of Portuguese syntax in LFG-PFM 49
3.1.2 Basics of Portuguese morphology in LFG-PFM 58
3.2 Further aspects of LFG 66
3.2.1 Semantic structure 66
3.2.2 LFG and diachrony 68
3.3 Extending PFM 71

3.3.1 Taylor's (2012) adaptations of PFM ... 71
3.3.2 Generalized Paradigm Function Morphology (Spencer 2013) ... 79
3.3.3 PFM2 (Stump 2016) ... 83
3.3.4 Periphrasis in PFM (Biswas 2019) ... 88
3.4 Chapter summary ... 94
Chapter 4 Portuguese double past participles in synchrony ... 95
4.1 Past participles and their properties ... 96
4.1.1 Functions of past participles in Portuguese ... 96
4.1.2 Forms of past participles in Portuguese ... 102
4.2 Lexical category of past participles ... 104
4.2.1 Differentiation of verbs and adjectives ... 104
4.2.2 Distinguishing the lexical category of Portuguese past participles ... 108
4.3 The double past participle phenomenon ... 115
4.3.1 Double past participles from a cross-linguistic perspective ... 117
4.3.2 Double past participles in Portuguese ... 119
4.3.3 Previous approaches to Portuguese double past participles ... 125
4.3.4 Interim summary ... 133
4.4 A formal LFG-PFM account of double past participles ... 135
4.4.1 Morphological and syntactic views of double past participles ... 136
4.4.2 Theoretical approaches to the lexical category of past participles ... 139
4.4.3 The place of past participles in the verbal paradigm ... 147
4.4.4 Lexical specification in the double past participle phenomenon ... 151
4.4.5 PFM rules for past participles as complex stems ... 155
4.4.6 Worked examples ... 163
4.4.7 Number and gender agreement of the passive past participle ... 170
4.4.8 Capturing syncretism between the two past participles ... 173
4.4.9 Accounting for past participles in periphrases ... 176
4.4.10 An integrated LFG-PFM account of the passive periphrasis ... 181
4.5 Conclusions ... 186
Chapter 5 Portuguese double past participles in diachrony ... 188
5.1 Diachronic trajectories of double past participle forms ... 190
5.1.1 Methodological questions in diachronic corpus work ... 191
5.1.2 Past participles in Latin ... 193
5.1.3 Double past participle forms from Latin to Portuguese ... 196
5.1.4 Attestations of different forms in the history of Portuguese ... 205

5.1.5 Divergence across varieties: double past participles in Brazilian Portuguese 219
5.1.6 Summary of diachronic trajectories 226
5.2 Analysing diachronic change 229
5.2.1 The role of major inflection classes in the diachrony of the past participles 229
5.2.2 Lexical specification in an LFG-PFM model of change 232
5.2.3 Analogical change in the diachrony of Portuguese past participles 234
5.3 PFM rules to account for diachrony 237
5.3.1 Stem roots 238
5.3.2 Stem formation rules and inflection classes 244
5.3.3 Case studies: PFM rules for Brazilian Portuguese varieties 250
5.3.4 Cross-paradigm diachronic processes: reanalysis of paradigm membership 261
5.3.5 Integration with LFG: Lexemic entries 266
5.4 Conclusions 272
Chapter 6 Portuguese grammaticalised pronouns in diachrony 274
6.1 Introduction to Portuguese pronouns and agreement mismatches 275
6.1.1 Pronouns in Portuguese 275
6.1.2 Agreement: mismatches, imposters and hybrids 283
6.1.3 Beyond semantic and syntactic agreement: INDEX and CONCORD 288
6.2 Use of grammaticalised pronouns in Portuguese 291
6.2.1 Data: second person você and vocês 292
6.2.2 Data: first person plural a gente 309
6.2.3 Interim summary 319
6.3 A formal LFG-PFM account of grammaticalised pronouns 320
6.3.1 Relationship between nouns and pronouns in the formal framework 321
6.3.2 PFM analysis of nouns and pronouns 330
6.3.3 Mismatches between syntax and semantics and their representation in LFG 337
6.3.4 Formal representation of grammatical features 340
6.4 Worked formal representations 343
6.4.1 Formal representations of second person singular pronouns 348
6.4.2 Formal representations of second person plural pronouns 354
6.4.3 Formal representations of first person plural pronouns 358
6.4.4 Lexical and lexemic entries for inherited and grammaticalised pronouns 365
6.4.5 PFM rules for pronouns 370
6.5 Analysis: LFG-PFM insights into grammaticalised pronouns 372
6.5.1 Adaptations to LFG representations 372

6.5.2 Value of an LFG-PFM approach to mismatches 373
6.5.3 Questions raised and areas for future research 375
6.6 Conclusions 379
Chapter 7 Conclusions from the thesis 382
7.1 Key contributions of the thesis 382
7.2 Directions for future research 386
Appendices 390
Appendix I Verbal paradigms: simplex forms 391
Appendix II Periphrastic or compound verb forms 396
Appendix III PFM rules and lexical data for Portuguese 399
Appendix IV PFM rule blocks for Brazilian Portuguese past participles 409
References 417

List of Figures

Figure 1 Map of the Portuguese-speaking world 10
Figure 2 The LFG architecture 20
Figure 3 The set of lexical entries $\mathscr{L}$ and its requirements 33
Figure 4 Schematic representation of PFM2 architecture 84
Figure 5 Past participles in the verbal paradigm 147
Figure 6 Partial paradigm of the regular verb *cantar* 'sing' 149
Figure 7 Partial paradigm of the double past participle verb *acender* 'light, turn on' 150
Figure 8 Past participles in the verbal paradigm 178
Figure 9 Relative frequencies of irregular and regular past participle forms for first conjugation verbs with auxiliary *ter* 'have' 208
Figure 10 Relative frequencies of irregular and regular past participle forms for second conjugation verbs with auxiliary *ter* 'have' 211
Figure 11 Relative frequencies of irregular and regular past participle forms for first conjugation verbs with auxiliary *ser* 'be' 213
Figure 12 Relative frequencies of irregular and regular past participle forms for second conjugation verbs with auxiliary *ser* 'be' 216
Figure 13 Possible trajectories of Portuguese verbs' past participle forms 227
Figure 14 Portuguese verbs exemplifying possible trajectories of the past participle forms 227
Figure 15 Possible outcomes of grammaticalisation process for pronouns 338
Figure 16 Analysis of 'Kim ate at noon' 346

List of Tables

Table 1 Key dates in the Portuguese Discoveries 8
Table 2 Major differences between Brazilian and European Portuguese 11
Table 3 Periodisation of Portuguese 12
Table 4 Tenses and moods of Portuguese verbs 59
Table 5 Periphrastic forms with *ter* 'have' and *ser* 'be' 63
Table 6 Stages of lexical specification in the transition from pronoun to agreement 70
Table 7 Types of lexical relatedness 81
Table 8 Paradigms in PFM2 and their cells 83
Table 9 Portuguese past participles by verb type 102
Table 10 Morphological patterns of irregular past participles 103
Table 11 Typical properties of verbs and adjectives 107
Table 12 Lexical category properties of verbs and adjectives in Portuguese 109
Table 13 Number and gender agreement in Portuguese 110
Table 14 Category properties of past participles 114
Table 15 Terminology for past participle forms 121
Table 16 Extension of athematic past participle forms 130
Table 17 Forms and functions of the past participle for different types of verbs 134
Table 18 Summary of morphological patterns of the past participle 157
Table 19 Number and gender agreement in Portuguese 170
Table 20 Examples of passive past participles with number and gender agreement 170
Table 21 Development of Portuguese verb conjugations from Latin 194
Table 22 Past participles across Latin conjugations 195
Table 23 Sound changes in Latin participles 195
Table 24 Trajectories of past participles of first conjugation verbs. 201
Table 25 Trajectories of past participles of second and third conjugation verbs 204
Table 26 Frequencies of selected Portuguese verbs 205
Table 27 Frequencies of past participle forms of verbs in the first conjugation with *ter* 'have' (*Corpus do Português* Genre/Historical) 207
Table 28 Frequencies of past participle forms of verbs in the second conjugation with *ter* 'have' (*Corpus do Português* Genre/Historical) 210
Table 29 Frequencies of past participle forms of verbs in the first conjugation with *ser* 'be' (*Corpus do Português* Genre/Historical) 212
Table 30 Frequencies of past participle forms of verbs in the second conjugation with *ser* 'be' (*Corpus do Português* Genre/Historical) 215
Table 31 Past participle forms included in the *Dicionário do Português Medieval* 218
Table 32 Frequencies of participle form across dialects in 20th century data from *Corpus do Português* 220
Table 33 Basic stem roots of modern Portuguese verbs 240
Table 34 Pronoun forms in Portuguese: full paradigms 277
Table 35 Second person singular forms of address 293
Table 36 Second person plural forms of address 295
Table 37 Sociolinguistic implications of second person singular forms of address 296
Table 38 Sociolinguistic implications of second person plural forms of address 297
Table 39 Features of *vossa mercê* 306
Table 40 Features of *você* in European Portuguese 306

Table 41 Features of *você* in a less grammaticalised variety of Brazilian Portuguese 306
Table 42 Features of *você* in a more grammaticalised variety of Brazilian Portuguese 307
Table 43 Features of *vocês* in a variety with more syntactic agreement 309
Table 44 Features of *vocês* in a variety with mixed agreement 309
Table 45 Features of *a gente*: original noun phrase, with syntactic agreement 319
Table 46 Features of *a gente*: less grammaticalised, with mixed agreement 319
Table 47 Features of *a gente*: more grammaticalised, with semantic agreement..... 319
Table 48 Pronoun forms in Portuguese: full paradigms 334
Table 49 Four distinct sets of features to analyse mismatches 342
Table 50 Periphrastic forms with a single auxiliary verb..... 396
Table 51 Periphrastic forms with two auxiliary verbs 397
Table 52 Periphrastic forms with subjunctive auxiliary verb..... 397
Table 53 Periphrastic forms with subjunctive of *ter* 'have' auxiliary and *ser* 'be' auxiliary..... 398

List of Abbreviations

↑	refers to the f-structure of the immediately dominating c-structure node
ACC	Accusative
Adj	Adjective
agr	Agreement
BSR	Stem base selection rule
C1	First conjugation class
C2	Second conjugation class
C3	Third conjugation class
cat	Category
COND	Conditional
c-structure	Constituent structure
$\mathcal{D}$	Description function
DEF	Definite
Det	Determiner
DM	Distributed Morphology
DOBJ	Direct object
DoublePTCP	Double past participle
DP	Determiner phrase
f-description	Functional description
f-structure	Functional structure
FUT	Future
GEND	Gender
GER	Gerund
gerImp	Imperative based on gerund

HPSG	Head-Driven Phrase Structure Grammar
i	variant index
IFD	Identity Function Default
IMPF	Imperfect
ind	Indicative
INDOBJ	Indirect object
INF	Infinitive
IPFV	Imperfective
IRREGPTCP	Irregular past participle
IrrImp	Irregular imperative
i-structure	Information structure
$\mathscr{L}$	Set of Lexical Entries
LE	Lexemic Entry
LFG	Lexical-Functional Grammar
LI	Lexemic Index
md	Mood
m-entry	Morphological entry
N	Noun
Nar	Narrowest applicable rule
NEG	Negation
NOM	Nominative
NUM	Number
past 1	Imperfect past tense
past 2	Simple past tense (preterite)
PF	Paradigm function

PF (in DM)	Phonetic or phonological form
PFM	Paradigm Function Morphology
p-form	Prosodic or phonological form
PL	Plural
plup	Pluperfect
pplImp	Imperative based on past participle
PRES	Present tense
p-string	Prosodic string
p-structure	Prosodic or phonological structure
PTCP	Past participle
PYTA	pretérito y tiempos afines ('perfect and related tenses')
$\mathscr{R}$	Morphological realisation relation
REFL	Reflexive
REGPTCP	Regular past participle
RR	realisation rule
SBJV	Subjunctive
SF	stem formation function
s-form	Syntactic form
SG	Singular
S-IrregPTCP	Irregular past participle with /s/
SR	Stem formation rule
SSR	Stem selection rule
s-string	Syntactic string
s-structure	Semantic structure
SUBJ	Subject

subj	Subjunctive
TAM	tense, aspect and mood
T-IrregPTCP	Irregular past participle with /t/
tns	Tense
TruncPTCP	Truncated irregular past participle
V	Verb
β	Function from prosodic string to prosodic structure
ι (in LFG)	Function from semantic structure to information structure
ι (in PFM)	Morphomic index
π	Function from syntactic string to constituent structure
ϱ	morphosyntactic property index
σ (in LFG)	Function from functional structure to semantic structure
σ (in PFM)	Set of morphosyntactic features
τ (in PFM)	Morphosyntactic property index
τ (in PFM2)	Set of morphosyntactic features associated with a stem
φ	Function from constituent structure to functional structure
χ	array of phonetic strings (typically input to PF or SF)
ψ	array of phonetic strings (typically output of SF)
1	First person
2	Second person
3	Third person

Glosses generally follow the Leipzig Glossing Rules, with additions as required. However, since the hyphen is used in standard Portuguese orthography to separate postverbal clitic pronouns from verbs, I use the hyphen instead of an equals sign in this context, and do not use a hyphen to indicate other putative morpheme boundaries.

Abstract

This thesis proposes an integration of Paradigm Function Morphology (PFM) as the morphological component within the Lexical-Functional Grammar (LFG) architecture, applying the combined theoretical framework LFG-PFM to account for diachronic data. The theory is developed by application to two case studies from Portuguese involving mismatches between different levels of grammatical structure and resulting in divergence between varieties of the language. In each case, I give a new presentation of the Portuguese phenomena, which have not previously been treated in LFG or PFM.

The first case study is the double past participle phenomenon, in which an arbitrary set of verbs unexpectedly have two past participle forms split between active and passive contexts, where most verbs only have one form. There is significant variability in the application of the phenomenon to lexemes over time and across varieties of Portuguese. I show that this can be successfully analysed in LFG-PFM using lexical specification and inflection classes. In addition, I clarify the role of irregular stem roots in the double past participles and identify possible trajectories of change. The second case study is Portuguese grammaticalised pronouns, which have developed from noun phrases and show interesting mismatch behaviour in their agreement patterns. My analysis uses distinctions in the PERSON feature at the levels of semantic structure and functional structure to capture various mismatch patterns seen across varieties of European and Brazilian Portuguese, including differences in the choice of coreferent pronouns and verb agreement.

The two case studies together demonstrate the success of the LFG-PFM approach of treating separate modules of the grammar independently. This allows a fine-grained account of the incremental changes that contribute to the development of distinct varieties of a language, and offers new insights into the morphological and syntactic phenomena at issue.

Acknowledgements

I gratefully acknowledge the many people who have supported me throughout the process of producing this thesis. In particular, I thank my supervisor Louise Mycock, who has made completing the DPhil more enjoyable and rewarding than I thought possible. I am so grateful for your guidance, feedback, positivity, and encyclopaedic knowledge of LFG.

Being part of Oxford's Faculty of Linguistics, Philology and Phonetics has meant I have benefited from the expertise of a whole community of syntacticians, as well as many specialists in the Romance languages. For education in the ways of LFG I would like to thank Mary Dalrymple, Ash Asudeh, Jamie Findlay, Alex Biswas, Stephen Jones, and the Syntax Working Group in its many guises. On the Romance side, I am particularly grateful to Sandra Paoli, who supervised my MPhil work on double past participles, Martin Maiden, Víctor Acedo-Matellán, Xavier Bach, Brittany Hause, and other members of the Discussing Romance group. I would like to extend further thanks to my examiners, Mary Dalrymple and Louise Esher, for their illuminating and constructive comments on my work during the viva, which have much improved the thesis.

Thanks are also due to the many friends I have made in Linguistics, especially those from my MPhil and DPhil cohorts: Hannah Davidson, Emily Lindsay-Smith, Katie Truslove and Chenzi Xu. It has been a pleasure to walk this path with you.

Since coming to Oxford to read Beginners' Portuguese and Linguistics as an undergraduate, I have been fortunate to be in close contact with the Portuguese sub-Faculty, and I would like to thank in particular Claire Williams and Claúdia Pazos-Alonso. I am grateful for the advice of Stephen Parkinson who first introduced me to Portuguese Linguistics and encouraged me to start carrying out my own research.

While writing this thesis I have benefited from the support of the Oxford Writing Programme. I owe many thanks to Emily Troscianko, the programme leader, as well as to my writing partners,

Kim Webb, Rowan Anderson and Hannah Davidson, for helping make all 100,000 words happen. Special mentions also go to Jane Ellard, Ellie Powell and Emilie de Bree, to my wonderful proof-readers, and to my Merton friends.

To Caity Ross (aka Catherine Taylor), whose analysis of Romance verbal morphology has played a key role in the development of my thesis: thank you for answering my questions so generously. I am also grateful to the many linguists who have commented on or asked questions about my work at conferences and seminars, including at ICHL24, LFG19, the Cambridge Romance Linguistics Seminar 2020, and LFG20.

This work, and attendance at conferences during my studies, has been financially supported by a studentship from the AHRC and the Birla Studentship, as well as travel grants from Trinity College and the Faculty of Linguistics. I gratefully acknowledge this support, without which the work would not have been possible.

Finally, I would like to thank my family for inspiring and encouraging me: my parents, Michèle and Robert Thomas, my brother Edward, my grandmother Rosemary, and my partner Stephen Thatcher. Thank you for your love and support, for which I will always be grateful.

Chapter 1 Introduction

This thesis applies the Lexical Functional Grammar (LFG) architecture, together with Paradigm Function Morphology (PFM) as the morphological component, to provide novel analyses of Portuguese morphology and syntax. It has two aims: to contribute to the development of a combined LFG-PFM theory that can account for phenomena involving both syntax and morphology, and to offer new insights into syntactic and morphological aspects of Portuguese. I focus on two case studies from Portuguese. The first is the double past participle phenomenon, where some verbs have two morphologically distinct past participles rather than the expected one form. The second is grammaticalised pronouns, where noun phrases come to function as pronouns and take on, to a greater or lesser degree, the grammatical features expected of pronouns.

My analysis also extends to diachronic aspects of the phenomena studied here, demonstrating that the LFG-PFM theory can be applied to account for language change and the divergence of language varieties. I show that the LFG-PFM approach allows powerful and detailed analyses of the small changes across morphology and syntax that contribute to the development of distinct linguistic varieties, which is evident in the variation seen across the case study phenomena between Portuguese varieties at a large or small scale, between and within Brazilian and European Portuguese.

The case studies represented here, the double past participle phenomenon and grammaticalised pronouns, reflect different aspects of the grammar. The first focuses more on morphology and particularly verbs, while the second is more syntactic in nature and relates to the noun phrase. In each case, we see divergence from the expected alignment of features across levels of linguistic structure, resulting in mismatch behaviour. I show that combining the parallel architecture of LFG with the powerful inflectional morphological component of PFM enables fine-grained accounts of these mismatches across levels of grammatical representation. In my accounts of the two case

studies, I show that the parallel architecture enables clear insights into mismatches that span morphology, syntactic structure and semantic structure.

1.1 Research questions

This thesis contributes to answering two types of questions: the first type relates to the theoretical framework of LFG and its combination with PFM, and the second relates to the Portuguese case studies in morphology and syntax.

On the theoretical side, this thesis provides a joined-up account of morphological and syntactic phenomena within the LFG architecture. Recent work has proposed a detailed model of the interface with morphology (Dalrymple 2015; Dalrymple, Lowe, and Mycock 2019, chap. 12), but does not propose a model of morphology itself. By applying PFM as the morphological component with this interface to analyse particular linguistic phenomena, I contribute an answer to the question:

1. How exactly does combining LFG with PFM work?

A second theoretical question relates to the application of the LFG-PFM model to diachronic phenomena, as there is no existing work on applying PFM in this way. In particular, I focus on the use of LFG-PFM to model incremental language changes that result in the divergence of language varieties.

2. How can diachronic phenomena, including linguistic divergence, be analysed using LFG-PFM?

On the Portuguese side, this thesis contributes to filling gaps in existing accounts of Portuguese morphology and syntax by presenting new data about particular morphosyntactic phenomena that have never previously been treated from a LFG or PFM perspective. I provide new analyses of two phenomena: double past participles and grammaticalised pronouns.

3. What new insights does the LFG-PFM analysis provide into the double past participle phenomenon and grammaticalised pronouns in Portuguese?

Applying the LFG-PFM framework to analyse these phenomena reveals that some aspects of them, such as stem distributions in the double past participles, and mismatches in the person feature associated with grammaticalised pronouns, cannot be fully accounted for in standard LFG. I therefore also answer the question:

4. What extensions to the theory are needed to account for the Portuguese case studies?

In the next section, I give further details about the contributions of this thesis, considering novel aspects of the research and providing further information about the two case studies.

1.2 Novel aspects of this research

One of the key novel contributions of this research is to demonstrate a detailed application of PFM as the inflectional morphological component within the LFG architecture, according to the interface set out by Dalrymple (2015; Dalrymple, Lowe, and Mycock 2019). This interface develops previous work in the area of the interface between syntax and morphology in LFG (Butt, Niño, and Segond 1996; Sadler and Spencer 2001; Frank and Zaenen 2002). It provides an integrated approach to morphological and syntactic phenomena, including those where the morphology itself is particularly interesting.

Autonomous morphological phenomena, such as the arbitrary distribution of stems throughout a paradigm, have long been treated independently of LFG in traditional morphological approaches. The syntactic basis of LFG has meant that work on morphology in that theory has generally focused on the interface between morphology and syntax, or the morphological realisation of morphosyntactic features, rather than considering purely morphological phenomena (see for example Sadler and Spencer 2001; 2004; Sadler and Nordlinger 2018). Here I demonstrate the usefulness of integrating PFM as the morphological component in LFG, in order to provide a

more comprehensive analysis of purely morphological phenomena within the wider LFG architecture. This builds on existing work on the integration of PFM with LFG (Luís and Sadler 2003; Luís 2004; Luís and Otoguro 2005; Biswas 2017; 2019). The powerful inflectional morphological 'engine' of PFM allows a detailed account of inflection, providing pairs of fully inflected wordforms with morphosyntactic property sets. These are required for the production of a complete set of lexical entries in LFG, which play a crucial role in the parallel architecture. As I show in this thesis, particularly in Chapter 4, considering the morphological component alongside the lexical and lexemic entries, and together with other modules of the grammar (syntax and semantics), allows new insights into the mismatches that can exist between different levels of structure, as well as divergences found across different varieties and over time.

The other major theoretical contribution of the thesis is to extend existing understanding of the use of the LFG architecture to account for diachronic change, following previous work carried out in this direction (Butt and King 2001; Vincent 2001; Vincent and Börjars 2010; Börjars, Harries, and Vincent 2016). In addition, I demonstrate the potential of applying the PFM theory for the same purposes. There is a growing body of work in LFG that uses the modular approach to provide detailed accounts of diachronic developments, especially in areas such as the grammaticalisation of content items to functional items (Butt and Lahiri 2013; Börjars and Vincent 2017). These approaches emphasise the advantages of treating separately the different modules of the grammar, such as semantics and syntax. However, while the complexities of morphological change have long been recognised by morphologists, up to this point, these aspects have not been integrated in diachronic accounts based on the LFG architecture, with existing treatments focusing on different aspects of syntax and semantics, even in instances where morphological forms also undergo change (e.g. Vincent and Börjars 2010). I demonstrate that the PFM formalism, like LFG, is extremely useful for capturing and accounting for incremental distinctions in diachronic variation. Differences in the application of a small number of PFM rules can account for wider morphological changes that are visible in the history of a language, as we see in Chapter 5. I

propose here that this approach can lead to new insights into the types of morphological and morphosyntactic change that we expect to see. Some further suggestions for the continuation of this research are made in the conclusions to my work (Chapter 7).

The thesis also develops further aspects of the LFG architecture, including the components of functional structure and semantic structure that are needed to account for mismatch behaviour in synchrony. This is developed in my analysis of Portuguese grammaticalised pronouns (Chapter 6). Again, the proposals made relating to these modules of the grammar can be deployed for diachronic analysis too: they allow a clear insight into the incremental differences seen across divergent varieties of a language.

The primary aim of the thesis is theoretical: to implement and adapt as necessary the LFG-PFM theoretical framework in order to account for complex morphosyntactic phenomena in synchrony and in diachrony. The two theories are developed and strengthened by their application to case studies from Portuguese which have not previously been analysed from these theoretical perspectives, although work has been carried out on other areas of Portuguese morphology and syntax in PFM, LFG or both (Luís and Sadler 2003; Luís 2004; Luís and Otoguro 2004; 2005; Bonami and Luís 2015; C. Taylor 2012). The work therefore also fulfils a secondary aim of providing novel, formally detailed analyses of complex morphosyntactic phenomena from Portuguese which have attracted relatively little theoretical attention, and none from an LFG or PFM perspective.

The first case study of the thesis is the double past participle phenomenon, in which some verbs have two morphologically distinct past participles split between active and passive contexts, where the vast majority of verbs across Portuguese (and indeed across the Romance languages more generally) have only one form. The double past participle phenomenon has been studied in some detail from an empirical or variationist perspective (e.g. Chagas de Souza 2007; Schwenter et al. 2019), and theoretically within Distributed Morphology (e.g. Lobato 1999; Nevins and Rodrigues

2014), but not within PFM, despite the existence of previous PFM analyses of parts of the verbal inflection system of Portuguese (Luís 2004; C. Taylor 2012). Similarly, the grammaticalised pronouns present unusual mismatch properties regarding the person feature, and have received considerable empirical attention, particularly focusing on their sociolinguistic properties (e.g. C. R. dos S. Lopes 2003; Thomé-Williams 2004; Lara Bermejo and Guilherme 2018), but there is no existing formal theoretical analysis in LFG, although analyses of some aspects of their grammatical behaviour exist in other frameworks (C. R. dos S. Lopes 2003; M. Taylor 2009; Costa and Pereira 2013; Marcotulio, Vianna, and Lopes 2013). My thesis presents the first LFG-PFM analyses of these two case studies, one that is more morphological and the other more syntactic, one relating to verbs and the other relating to the noun phrase. This adds to our knowledge and understanding of Portuguese morphosyntax, and provides an improved theoretical framework for future work on similar phenomena in different language varieties in Romance and beyond.

More broadly, the thesis demonstrates that there is considerable potential in combining LFG with PFM for accounting for wide-ranging morphological and syntactic change, in Portuguese specifically, but also in other languages. In Chapter 7 I provide suggestions for the extension of this research programme, with a particular emphasis on a cluster of interrelated changes in the morphology and syntax of Portuguese that includes the replacement of inherited pronouns by grammaticalised pronouns, the reduction of person/number distinctions in the paradigm, and changes in the use of null versus overt subjects. The LFG-PFM framework provides an ideal architecture for the examination of such interrelated changes, given that it takes morphology and syntax as independent but closely related modules in the grammar, allowing an account of mismatches and interactions between the two. I also propose that the framework could be fruitfully applied to investigate matters such as the typology of changes in inflectional morphology and their relationship to syntactic properties, within Portuguese and cross-linguistically.

1.3 Brief introduction to Portuguese

In this section, I give a brief introduction to Portuguese, including the history of the language and its division into modern varieties, to provide context for the rest of the thesis.

Portuguese is an Ibero-Romance language, estimated to be spoken by over 230 million people (Segura 2013, 71). It developed as a direct continuation of the Latin spoken in western Iberia (Bechara 2015, 25), following the arrival of the Romans in the peninsula from around 200BC (Castro 2006). As noted by Ilari (2013, 50), the ancestor language of Portuguese was not literary Latin, but a spoken variety brought to Iberia by soldiers, traders and settlers. This spoken variety of Latin underwent many developments over the centuries, and was affected by other languages brought to the region, including by Germanic tribes like the Visigoths in the 400s (Bechara 2015, 25). Although most of the Iberian peninsula was occupied by the Muslims in 711, modern-day Galicia and northern Portugal were not, allowing a continuous development of the language. Throughout the Reconquest of the Iberian peninsula by Christian forces, which began in 718 and lasted until 1249, the language spoken in the north spread southwards to the Algarve (southern region of Portugal), replacing the Arabic-influenced Romance varieties, known as Mozarabic, that had become widely spoken in those areas. Portugal became a kingdom in 1139 (Bechara 2015, 26), contributing to a growing sense of a distinct people with a distinct language. The emergence of the written language in the thirteenth century, albeit with substantial geographical variation, led to a strong literary tradition (Dubert and Galves 2016, 411).

The language spoken in the west of Iberia from medieval times until the Renaissance is sometimes known as Galician-Portuguese (Dubert and Galves 2016, 412). From around 1400, Portuguese and Galician are generally treated separately, with the river Minho treated as a linguistic border (Bechara 2015, 26).

The gradual spread of the language southwards is argued to have contributed to modern regional variation within Portuguese. Castro (2006) argues that modern-day northern varieties (Northern

European Portuguese, as well as Galician) retain conservative features, being the original areas in which Galician-Portuguese was spoken, while the Central-Southern European Portuguese varieties are more innovative, because their history in that location only dates back to the Reconquest.

As a result of Portuguese expansionist policies beginning in the 1400s, Portuguese was also taken overseas, including to Brazil and colonies in Africa and Asia. The following table, based on information from Castro (2006, 150–51) summarises some important dates in these Discoveries.

1415	Traditional date of start of Discoveries. Conquest of Ceuta in modern-day Morocco
1418	Discovery of Porto Santo (then uninhabited Atlantic island near Madeira)
1425	Discovery of Madeira (then uninhabited)
1427	Discovery of the Azores (then uninhabited)
From 1469	Exploration of Gulf of Guinea and crossing of the Equator; discovery of São Tomé and Príncipe
1488	Rounding of Cape of Good Hope
1498	Contact made with Calicut, West India
1500	Arrival of Portuguese fleet in Brazil
Early 1500s	Further exploration and trading in Asia, including establishing trading ports in Goa, Malacca, Macau and Nagasaki

Table 1 Key dates in the Portuguese Discoveries

Brazil was colonised from the 1500s, but use of the Portuguese language was not extensive until the mid 1700s: before that time, it coexisted with a lingua franca known as *língua geral* 'general language', which was based on the indigenous language Tupi. The Portuguese spoken in Brazil was influenced by this lingua franca and by the languages of African slaves brought to Brazil by the Portuguese (Baxter 1991, 19). After centuries as a colony, Brazil gained independence from Portugal in 1822, contributing to the emergence of a national identity. After this point written Brazilian Portuguese began to slowly incorporate more features of the language as spoken in Brazil,

with a growing recognition in the 1800s and 1900s that Brazilian Portuguese represented an independent variety (Baxter 1991, 21). Throughout its history, the development of Brazilian Portuguese has been strongly influenced by the European Portuguese norm, particularly in the written language, and significant linguistic prejudice is associated with speaking varieties further from the standard (O'Neill and Massini-Cagliari 2019). However, there is also an increasingly wide recognition of Brazilian Portuguese as a separate norm, leading to the analysis of Portuguese as a pluricentric language, that is, a language with more than one standard variety (Baxter 1991).

Outside Portugal and Brazil, varieties of Portuguese are spoken in many areas, including in Africa (Angola, Cape Verde, Guinea-Bissau, Mozambique, and São Tomé and Príncipe) as well as East Timor and Macau. Portuguese-based creoles are also spoken along areas of the coast of India and in some parts of Southeast Asia, as well as in a small number of locations in America outside Brazil (Segura 2013). Areas where Portuguese has a recognised language status are marked on the map (Figure 1).



Figure 1 Map of the Portuguese-speaking world

In this thesis, I focus on European and Brazilian Portuguese, as the two major varieties of the language. There are high-level differences between European and Brazilian Portuguese including in the lexicon, phonology, morphology and syntax. As well as general processes of linguistic divergence, the rich context of language contact in Brazil has led to widespread grammatical change, some of which is the result of irregular linguistic transmission (Lucchesi 2001; 2017). This analysis has led to claims that Brazilian Portuguese is a semi-creole, or has some creole-like characteristics (see for example Tarallo 1993; Holm 1994; Baxter 1997; Parkvall 2000).

Some of the major morphological and syntactic differences between Brazilian and European Portuguese are summarised below, with references. Further discussion of these differences can be found in Galves and Kroch (2016) and Martins (2016).

Morphological or syntactic phenomenon	**Selected references**
Placement of clitic objects: tendency towards post-verbal (enclisis) in EP vs. towards pre-verbal (proclisis) in BP	(Pagotto 1993; Nunes 1993; Martins 2002; Scherre and Duarte 2016)
Realisation of direct objects: loss of clitics in favour of null objects or use of tonic pronouns with object function in BP	(Cyrino 1993; Nunes 1993; Cyrino and Matos 2016)
More frequent expression of subject pronouns i.e. loss of null subject parameter in BP	(M. E. L. Duarte 1993; M. E. L. Duarte and Silva 2016; M. E. L. Duarte 2017)
Replacement of inherited pronouns by grammaticalised pronouns for 1PL (*a gente*), 2SG(*você*) and 2PL (*vocês*), further advanced in BP than in EP	(C. R. dos S. Lopes 2003; Costa and Pereira 2013; C. R. dos S. Lopes and Brocardo 2016; Scherre and Duarte 2016)
Loss of person/number distinctions in the inflectional paradigm of verbs, in some BP varieties	(Vieira and Brandão 2017)
Less frequent marking of number/gender agreement in the nominal phrase, in some BP varieties	(Galves 1993)

Table 2 Major differences between Brazilian and European Portuguese

Having seen some of the large-scale differences between European and Brazilian Portuguese, it is now worth considering the distinctions that are drawn between narrower varieties. Modern varieties of European Portuguese are typically distinguished along geographical lines, with a wide distinction between mainland and island varieties, and more fine-grained distinctions between the different islands (Madeira and the Azores) and within Portugal between northern and central/southern varieties, or even at a smaller regional level (Raposo et al. 2013, vol. 1, chap. 5).

By contrast, within Brazilian Portuguese, geographical variation is generally treated as less significant than the distinction between 'popular' or 'vernacular' varieties and an idealised 'educated norm', although there is undoubtedly significant variation across the country, as can be straightforwardly seen in the Portuguese spoken in the major cities of Rio de Janeiro, São Paulo and Recife (Raposo et al. 2013, vol. 1, chap. 6). These different axes of variation are considered to some extent in parts of this thesis.

As well as considering divergence and variation at a synchronic level, we can also distinguish between historical periods of Portuguese. The historical development of Portuguese is typically divided into periods similar to those shown in Table 3 (adapted from Castro 2006, 73), with caveats about the arbitrariness of any such periodisation, given the gradual nature of language change.

1200 to 1420:	*português antigo*	Old Portuguese
1420 to 1550:	*português médio*	Middle Portuguese
1550 to 1800:	*português clássico*	Classical Portuguese
1800 onwards:	*português moderno*	Modern Portuguese

Table 3 Periodisation of Portuguese

Having this background information on Portuguese in place, I now move on to set out the structure of the rest of the thesis.

1.4 Structure of the thesis

The rest of the thesis is structured as follows. Chapter 2 includes the basic theoretical background for the thesis, setting out the main principles of LFG and PFM, as well as presenting the approach to the syntax-morphology interface used here that allows the two theories to be integrated. This chapter presents the different elements of this interface, including key concepts such as the lexical and lexemic entry. Chapter 3 continues this theoretical perspective, with a particular focus on the use of the two theories for the analysis of Portuguese. The chapter includes further details on

mismatches, demonstrating LFG accounts of mismatches from Portuguese involving several modules of the grammar. It also presents further theoretical background on LFG and PFM relevant to the work carried out here, including semantic structure in LFG, the use of LFG to study diachronic phenomena, and several proposed extensions to basic PFM theory. At the end of Chapter 3, the theoretical elements are in place for a detailed examination of the Portuguese case studies in the rest of the thesis.

The first of these case studies is the double past participle phenomenon, which is analysed from a synchronic perspective in Chapter 4 and from a diachronic perspective in Chapter 5. In the double past participle phenomenon, we see that while most Portuguese verbs have only one past participle form used across both passive and active periphrastic contexts, some verbs have two forms split between the two contexts. This discrepancy between the vast majority of Portuguese verbs, and the small number of double past participle verbs, can be interpreted as a mismatch in relation to the number of expected forms (one) and the number of forms found in double past participle verbs (two). Chapter 4 begins with a detailed examination of the functions and forms of past participles in Portuguese (section 4.1), as well as their lexical category properties (section 4.2), which in some instances are closer to adjectives than to verbs, allowing their analysis as 'mismatched categories' (Lowe 2019) or 'transpositions', in the sense of Spencer (2013). Section 4.3 presents information on the double past participle phenomenon in Portuguese, before section 4.4 gives my LFG-PFM analysis of the phenomenon. In particular, I give a new PFM analysis of the morphology of these forms, which allows both regular and irregular past participles to be formed correctly, and associated with the correct morphosyntactic property sets. I also show that the fine-grained lexicon of LFG-PFM allows an appropriate account of lexical specification which plays a key role in the phenomenon. There is substantial variation across varieties of Portuguese, which can be straightforwardly captured in an analysis that recognises the lexical arbitrariness of whether a verb has one or two past participles.

Chapter 5 builds on the formal LFG-PFM approach developed in Chapter 4 and extends it to the analysis of the double past participle phenomenon in diachrony. I present data from historical corpora of Portuguese, demonstrating that there is considerable variation in this respect across Portuguese over time (section 5.1). After considering the role of several different aspects of the theoretical model in accounting for these changes, including inflection classes and lexical specification (section 5.2), I present adaptations of the LFG-PFM rules and lexemic entries required to account for diachronic changes of various kinds (section 5.3): the loss of regular or irregular past participles, the gain of a new suppletive past participle, and the extension of an inflectional pattern in the 'innovative' past participles of Brazilian Portuguese. This chapter demonstrates that the LFG-PFM theory enables detailed representations of morphological and lexical change, including giving new insights into the types of changes that are more or less likely to occur. I show that although there are instances of irregular stem roots being lost from a verb's lexemic entry, it is rare for a new irregular stem root to be acquired (with one example of an irregular stem root being gained through suppletion). On the other hand, we see considerable flexibility in membership of the minor inflection classes that condition participation in the double past participle phenomenon, pointing towards the diachronic instability of exactly which verbs demonstrate it.

Chapter 6 presents the Portuguese grammaticalised pronouns, which have interesting mismatches in their grammatical features at the level of functional structure (a dimension of syntactic structure in the LFG framework), or between functional structure and semantic structure, visible in the properties marked on different agreement targets. For example, the second person singular pronoun *você* can in some varieties of Portuguese be used with second person singular coreferent pronouns, but third person verbs. This second case study contrasts with and complements the first: rather than involving verbs, it focuses on noun phrase phenomena, and where the first case study had more emphasis on morphological mismatch, the second is more syntactic and semantic. This allows a more extensive demonstration of the range of mismatch and divergence phenomena

that the LFG-PFM theory can describe and explain. The chapter gives an overview of pronouns in general in Portuguese, agreement and agreement mismatches (section 6.1), before presenting the grammaticalised pronouns in Portuguese (second person singular *você* and plural *vocês*, and first person plural *a gente*) and their agreement properties (section 6.2). Section 6.3 gives details of the formal analytical tools needed to account for these forms, and in section 6.4 I present a full LFG-PFM analysis of the diverse mismatch behaviour seen across the grammaticalised pronouns of different Portuguese varieties (European and Brazilian). I argue that to account fully for the grammatical behaviour of these forms, we must include PERSON as a feature in both INDEX and CONCORD, and that it is necessary to represent the NUMBER, PERSON and GENDER features in a detailed semantic structure, following proposals made by Liao (2010).

Finally, in Chapter 7, I bring together my key findings from the thesis, consolidating my argument in favour of taking an LFG-PFM approach to the analysis of morphosyntactic change from a diachronic perspective. I also suggest some directions for future research, setting out a programme for further study of morphosyntactic phenomena, including mismatches, using this theoretical framework.

Chapter 2 Theoretical background

This chapter introduces the theories of Lexical-Functional Grammar (LFG) and Paradigm Function Morphology (PFM). These two theories are used together in this thesis to provide detailed formal analyses of morphosyntactic phenomena in Portuguese which exemplify interesting patterns of mismatch and divergence, namely the double past participle phenomenon (Chapter 4 and Chapter 5) and the grammaticalised pronouns (Chapter 6).

The chapter is structured as follows. In section 2.1 I introduce the parallel architecture of LFG and some of its main properties, and set out the reasons for choosing this framework for the analysis carried out in this thesis. In section 2.2, I similarly present the crucial characteristics of PFM and describe the advantages of the theory for my purposes. Section 2.3 describes how PFM and LFG can be brought together, integrating PFM as the morphological component within the wider LFG architecture, including a consideration of the concepts of features and lexical categories as used in both theories. I give a detailed explanation of the syntax-morphology interface implemented here. Finally, in section 2.4, I bring together some of the main points made in this chapter.

2.1 Lexical-Functional Grammar

2.1.1 Fundamentals of LFG

LFG (Bresnan 1982; Kaplan and Bresnan 1982; Bresnan et al. 2015; Dalrymple, Lowe, and Mycock 2019) is a constraint-based theory of grammar, with distinct modules that capture different aspects of linguistic behaviour. These modules are separate but mutually constraining, and are related to each other through mapping functions. They are said to form a 'projection architecture', because the various different modules are projected from each other, though they exist simultaneously. Using LFG allows linguists to account for the different properties of particular languages, and

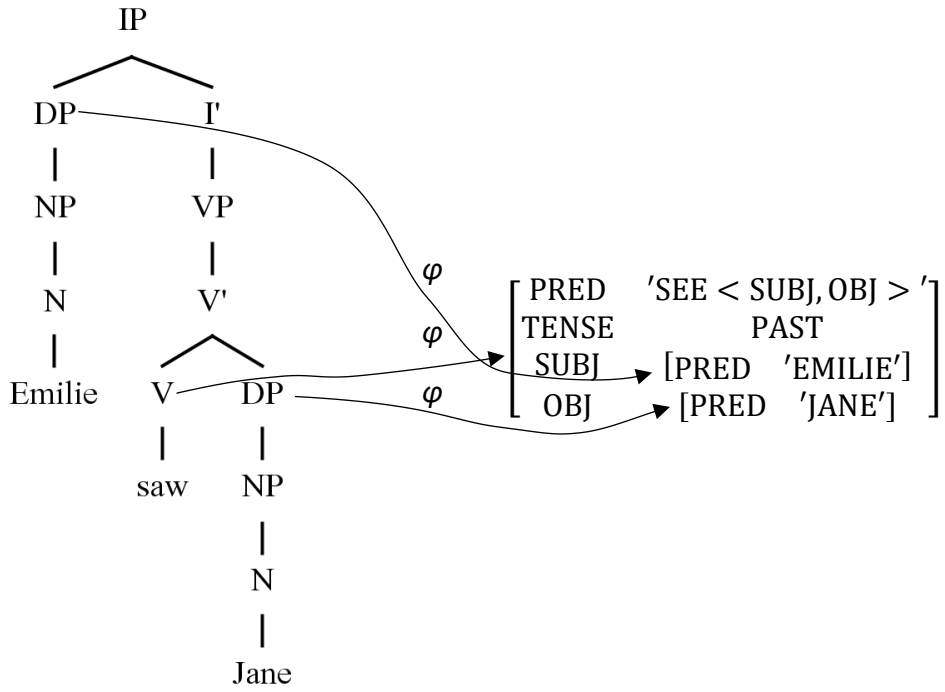
languages in general, using analyses appropriate for each type of information (Dalrymple, Lowe, and Mycock 2019, 265–67).

From a syntactic perspective, one of the most striking features of the LFG architecture is its use of two different levels of representation for syntactic phenomena. These are constituent structure (c-structure) and functional structure (f-structure). Functional structure represents the abstract functional features of sentences, and is notated in attribute-value matrices (AVMs), in which the ordering of particular properties is unimportant. On the other hand, constituent structure includes information about the structure of phrases, including constituency and linear order, and is represented in tree diagrams. LFG allows these two types of information to be treated separately, while also capturing the relationships that are regularly found between the two structures in any particular language. The two structures are related by the φ function, represented by an arrow, which connects corresponding parts of the c-structure and the f-structure. Since functional features are often shared across languages, while phrase structure and constituency vary greatly from language to language, we generally see cross-linguistic similarities at the level of functional structure, and more variation between languages at the level of constituent structure.

As an example, I present a basic c-structure and f-structure representation of the English sentence 'Emilie saw Jane' (1) and the Welsh sentence *Gwnaeth hi weld y draig* 'She saw a dragon' (2), (3). As the examples demonstrate, there are important differences in word order and constituency, but clear similarities at functional structure.

φ-mapping is shown for the most important nodes, that is for the DPs which map to the f-structure of the subject and object predicates, and for the V node which maps to the f-structure of the whole sentences. Following the principle of Economy of Expression (discussed in Dalrymple, Lowe, and Mycock 2019, 136–38), which promotes simpler and smaller c-structure trees as long as they obey well-formedness conditions, unnecessary bar level nodes are omitted from the c-structures.

(1) C-structure and f-structure representations of 'Emilie saw Jane' with simplified *φ*-mapping

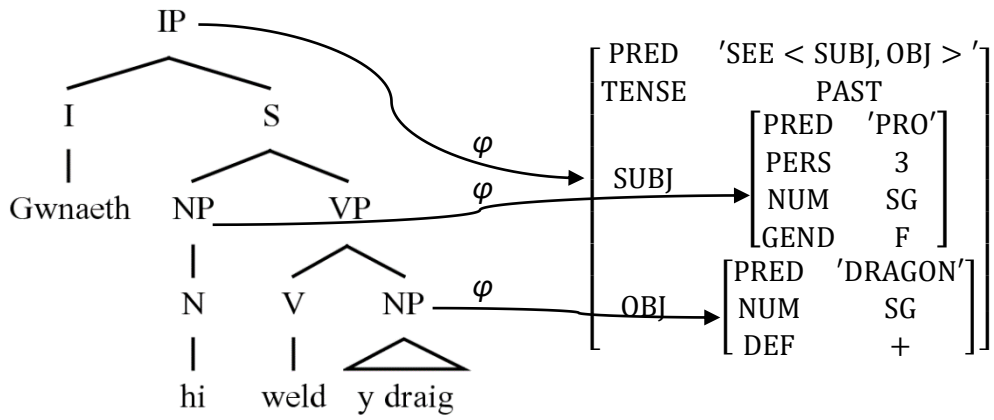


(2) Gwnaeth hi weld y draig

do.3SG.PST 3SG.F see the dragon

'She saw the dragon.'

(3) C-structure and f-structure for *Gwnaeth hi weld y draig* with simplified *φ*-mapping



C-structure originally proposed by Sadler (1997, 6), and adapted by Dalrymple, Lowe and Mycock (2019, 113).

Other levels of linguistic structure have also been proposed within the LFG architecture, alongside c-structure and f-structure, which have been the subject of most work. These further separate modules or levels of structure include semantics (represented at semantic structure or s-structure), information structure (i-structure), and prosodic or phonological structure (p-structure). An argument structure (a-structure) module is sometimes posited (Butt, Dalrymple, and Frank 1997), but this is not used in all LFG models (Asudeh and Giorgolo 2012). A further module that can be included in LFG work, which is most relevant to this thesis, is morphosyntactic structure (m-structure) (Butt, Fortmann, and Rohrer 1996; Butt et al. 1999; Frank and Zaenen 2002). However, following Dalrymple (2015; Dalrymple, Lowe, and Mycock 2019, 446), I do not assume a separate morphosyntactic structure, instead implementing Dalrymple's approach in which an independent morphological component provides fully inflected forms which are associated with morphosyntactic property sets, to give the full set of (correct) lexical entries. This reduces complexity within the architecture while still allowing a full account of morphology. Further details of these proposals for the interface between the morphological component and the syntax are

given in section 2.3. I also consider in that section the implications of integrating PFM as the morphological component.

The full LFG projection architecture, with no a-structure or m-structure, is presented in Figure 2, including the different modules and the mapping between them, represented as functions annotated with π, β, φ, etc. (Dalrymple, Lowe, and Mycock 2019, 471).

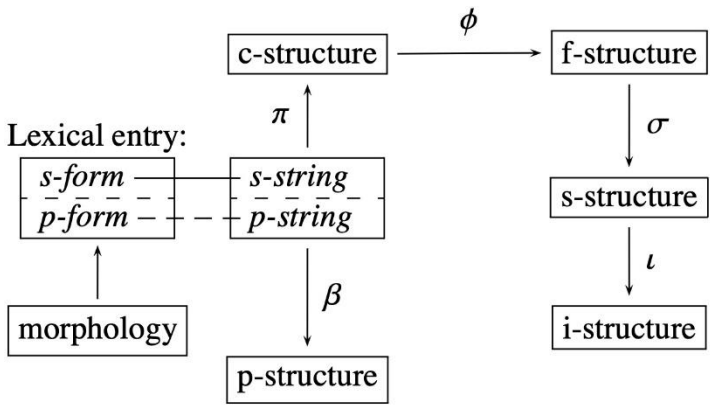


Figure 2 The LFG architecture, from *The Oxford Reference Guide to Lexical Functional Grammar*. © Mary Dalrymple, John J. Lowe, and Louise Mycock 2019. Reproduced with permission of Oxford Publishing Limited through PLSclear.

For my purposes, it is not necessary to present the details of the functions that map between all the different modules within LFG. However, the interaction of morphology with syntax, mediated by the lexical entry, is discussed in some depth in section 2.3. I also discuss the nature of semantic structure, and its relevance and use in my analysis of Portuguese, in Chapter 3.

LFG makes use of a rich lexicon, as suggested by the name of the theory. Many grammatical tendencies across lexical items are captured at the level of the lexicon, with both whole lexemes and individual wordforms having appropriate representations (lexemic entries and lexical entries, respectively). As described in more detail in section 2.3, the lexemic entry contains information relevant to the whole lexeme, including morphological information, an f-description (including semantic information), and the Lexemic Index. By contrast, the lexical entry relates the s-form (representing a form in the syntactic string), p-form (representing a form in the prosodic string,

presented using the IPA), category and f-description relevant to that particular wordform. By way of example, I give a sample lexemic entry for the verb 'see', and a sample lexical entry for its inflected past tense form 'saw'. I also provide a lexical entry for the proper noun 'Jane'.

(4) Lexemic entry for 'see'

LE < {M-ROOT: see, M-STEM1:saw, M-STEM2:seen, M-ICLASS:4, M- CAT:VERB},

{(↑PRED FN) = SEE, [**see**]},[1]

SEE1>

(5) Lexical entry for 'saw':

ℒ < saw, / sɔː/, V,

{(↑PRED)='SAW<SUBJ, OBJ>', (↑TENSE)=PAST, [**saw**]} >

(6) Lexical entry for 'Jane':

ℒ < Jane, / d͡ʒeɪn/, N,

{(↑PRED)='JANE', (↑NUM)=SG, [**Jane**]} >

Lexemic and lexical entries are crucial to capturing many morphosyntactic phenomena, and are discussed throughout this thesis, after being introduced fully in section 2.3.

Having seen some of the basic properties of LFG, I now move on to consider the advantages and justifications for applying the theory within this work.

[1] See section 2.3.1 for an explanation of the use of PRED FN here.

2.1.2 Advantages of LFG and justification for its use

As we have already seen, one of the major features of LFG is its architecture involving separate modules to account for different facets of the grammar. This is a particular advantage for work that seeks to deal with phenomena that involve more than one aspect of linguistic structure. For example, this work considers interactions between purely morphological patterns and syntactic properties. In LFG, it is straightforward to separate out distinct changes and represent their interactions, without implying that changes in one module necessarily bring about changes in others.

More generally, the strictly modular approach implemented in LFG provides flexibility to accommodate new analyses. This is particularly visible in modules outside syntax, where alternative approaches to non-syntactic phenomena can be formulated and directly compared, without disrupting other modules of the grammar. These different approaches may involve new proposals regarding the structure of the architecture, or involve the structure of one module in particular. To take an example which is not considered elsewhere in this thesis, various different LFG approaches have been taken to account for the contribution of prosody. There are different proposals for the location of prosodic structure (p-structure) in the architecture, either projected from c-structure (Mycock 2006) or from discourse structure (O'Connor 2006). Some approaches integrate the prosodic hierarchy (Mycock 2006) while others reject this (Bogel 2005).[2] The flexibility of the modular approach to the grammar offered by the LFG architecture can be contrasted with other approaches such as Minimalism (Chomsky 1995) and Head-Driven Phrase Structure Grammar (HPSG) (Pollard and Sag 1994) which use one level of analysis and one form of representation

[2] A fuller explanation of different approaches to prosody within LFG can be found in Dalrymple, Lowe and Mycock (2019, chap. 11).

(such as the phrase structure tree or the attribute-value matrix) to capture all different aspects of the grammar.

One focus of this thesis is mismatch phenomena. In analysing these phenomena, it is crucial to be able to consider different modules of the grammar separately. As noted by Francis and Michaelis (2003, 24), 'mismatches provide one of the most important rationales for distinguishing among different levels of grammatical representation', because they 'provide the contexts of dissociation which illuminate the modules or levels which otherwise yield isometric representations'. In LFG, mismatches are typically treated as 'incongruent mappings between relatively independent levels of linguistic structure, where incongruity is defined relative to a prototypical association of components from different levels' (Francis and Michaelis 2003, 6). An example is where a member of a syntactic category has semantic information typically associated with a different syntactic category, as discussed extensively in Francis and Michaelis's (2003) volume on mismatch. For example, a study by Yuasa and Francis (2003) included in that volume discusses Japanese subordination markers which have the formal properties of nouns and the semantic properties of subordinating conjunctions. In this thesis I discuss mismatches involving aspects of morphology, functional structure, and semantic structure.

The other major strand of this thesis considers the divergence of language varieties over time. An approach to grammar that includes multiple modules is relevant here, too. As languages diverge over time, we frequently see changes that affect some parts of the grammatical system, but which do not impact other modules at exactly the same time. This is frequently visible in grammaticalisation, where changes in a form's morphophonological properties may occur at a different rate from changes in its morphosyntactic properties. As a form moves from a more contentful role to a more functional role, it frequently undergoes phonological erosion and changes in syntactic category, alongside changes in grammatical features. Further, the same grammaticalisation pattern may take place slower or faster in particular varieties of one language.

In LFG, different types of changes can be analysed and accounted for separately (see for example Vincent 2001; Börjars, Harries, and Vincent 2016). The use of LFG to study diachronic phenomena is discussed in section 3.2.2.

Another area of particular relevance for this thesis is the treatment of morphology as separate from syntax in LFG. In some alternative approaches to syntax and the grammar, such as Distributed Morphology (DM) (Halle and Marantz 1993; 1994), which is often used in association with a Minimalist perspective, morphology is treated as a type of syntax that operates below the level of the word, with the same organising principles and representations. Although this approach promises theoretical efficiency, it does not recognise the unique properties of morphology that require separate representation and treatment. As has long been recognised by morphologists (Aronoff 1994; Maiden 2005), there are many autonomously morphological phenomena which cannot be accounted for in the same ways as syntax and which demand an appropriate treatment of a distinct kind. LFG is compatible with several different views of morphology, but recognises that morphological structure is distinct from syntax, which is particularly important for work that seeks to deal with autonomously morphological properties as well as their interaction with the syntax.[3]

Part of the rationale for the LFG perspective on morphology is the importance of the Lexical Integrity Principle, which has a long history in LFG (Dalrymple, Lowe, and Mycock 2019, sec. 4.4). According to this principle, 'every terminal node at the level of constituent structure corresponds to an individual word' (Dalrymple, Lowe, and Mycock 2019, 95–96), with individual words being formed in the morphological component rather than in the syntax. This follows from

[3] See Melchin, Asudeh and Siddiqi (2020) for an implementation of Distributed Morphology in an adapted LFG architecture.

the 'lexicalist hypothesis' that syntactic processes (originally, transformations) are not involved in the formation of words: instead, morphologically complete words are formed by lexical processes (Dalrymple, Lowe, and Mycock 2019, 135). From this perspective, a powerful morphological engine is needed to provide the inflected wordforms that correspond to terminal syntactic nodes. This role can be filled by a realisational morphological component like PFM, as discussed in sections 2.2 and 2.3. There are some challenges to the Lexical Integrity Principle where the mapping between morphological forms and terminal nodes is not one-to-one: examples include periphrastic verb forms like 'have seen', English auxiliary contractions like *I'll*, Portuguese preposition + article contractions like *do* < *de* + *o* 'of the' and *no* < *em* + *o* 'in the', and the position of clitic pronouns within inflected Portuguese verb forms like *comprá-lo-ei* 'I will buy it' (< *comprarei* + *o*). Some possible responses to these challenges, including the LFG concept of Lexical Sharing (Wescoat 2005) and Luís's treatment of Portuguese clitics (2004), are discussed in section 3.1.

To conclude this section, taking an LFG perspective presents a clear advantage for considering mismatch phenomena, particularly those that involve disruptions in the expected mapping between different levels of structure. The parallel architecture model of LFG allows a detailed analysis of attested mismatches, while also giving frequent insights into the types of mismatches that we might expect to find, and the types of divergence we might see between different linguistic varieties. An LFG analysis of these mismatches and variation takes each level of the grammar on its own terms, which cannot be offered by a derivational approach to syntax, such as Minimalist or other transformational approaches.

In the next section, I move on to describe the theory of morphology used here, PFM, and the reasons for this choice.

2.2 Paradigm Function Morphology

Many morphological phenomena cannot be described simply by a purely concatenative approach to morphology, such as 'Item and Arrangement' or 'Item and Process' models, in which

morphemes are added together to give an inflected form (see Robins 1959; C. Taylor 2012, chap. 2 for discussion). These include 'morphomic' patterns of stem selection where stems are shared between arbitrary or otherwise unmotivated sets of morphosyntactic properties. However, this type of complex morphological behaviour can be modelled effectively in Word and Paradigm models of morphology, of which PFM is an example. Word and Paradigm approaches more generally can provide clear insights into the diachrony of morphology, suggesting which types of changes are possible or likely (see for example Maiden 2018; Esher 2017; 2020).

2.2.1 Fundamentals of PFM

PFM (Stump 2001) is an inferential and realisational theory of morphology. 'Inferential' means here that the 'the systematic formal relations between a lexeme's root and the fully inflected wordforms constituting its paradigm are expressed by rules or formulas' (Stump 2001, 1), as opposed to lexical theories, including Distributed Morphology (Halle and Marantz 1993), where inflectional markers are listed in the lexicon. In a 'realisational' theory, 'a word's association with a particular set of morphosyntactic properties licenses the introduction of those properties' inflectional exponents' (Stump 2001, 2), as opposed to an 'incremental' theory, such as Articulated Morphology (Steele 1995), where the addition of inflectional exponents adds morphosyntactic properties to the meaning of the wordform.

PFM is rigorously defined, and designed to be formally explicit and computationally testable, as is seen in a computational implementation of the theory, the Paradigm Function Morphology Engine (PFME) (Finkel 2016). One of the key insights incorporated into PFM is that morphological paradigms are of primary importance in the inflectional system. A PFM analysis consists of a set of language-specific rules, presented in hierarchically organised rule blocks, that together define the morphological structure and content of the paradigm of a language (Stump 2001, 44).

There are substantial differences between the theory set out by Stump (2001), sometimes referred to as PFM1, and the newer PFM2 (Stump 2016) which sets out a more complex theory of

paradigms, with a separation between content paradigms, form paradigms and realisation paradigms. Some further details about PFM2 are given in section 3.3.3. Here, however, I focus on describing the basic insights of the original PFM, many of which are carried over to PFM2. I reserve the abbreviation PFM to refer to the theory set out in Stump (2001), and refer to the later theory as PFM2.

The heart of PFM is the Paradigm Function which, 'when applied to the root of a lexeme L paired with a set of morphosyntactic properties appropriate to L, determines the wordform occupying the corresponding cell in L's paradigm' (Stump 2001, 32). The function takes as its input a 'form/property-set pairing' of the form <X,σ> and determines the inflected form of the word that occupies the relevant cell in the paradigm, in a pairing with its properties (<Y,σ>). This is schematised below:

(7) $\langle X, \sigma \rangle \xrightarrow{\textit{Paradigm Function}} \langle Y, \sigma \rangle$

The Paradigm Function is expressed in terms of realisation rules, both rules of exponence, which 'directly specify the concrete exponents … associated with the property set being realized', and rules of referral, which 'refer the realization of some property set to some other realization rule(s)' (Stump 2001, 244). PFM also assumes the existence of two further classes of rules. The first is morphomic rules, which account for stem formation and indexing, 'defining the classes of stem-forms to which specific realization rules are defined as applying'.[4] The second is morphological meta-generalisations, which express the association that occurs 'when two or more realization rules

[4] The concept of 'morphomic' rules builds on the autonomously morphological phenomenon of morphomes, which are arbitrary relations between sets of forms and meanings. See section 2.2.2 for further details and examples.

are associated with the same morphophonological regularity' (Stump 2001, 245), allowing an account of regularities in stress or changes in vowel quality, as Stump demonstrates for Bulgarian (2001, 48–50). The morphomic rules and morphological meta-generalisations are not strictly part of the Paradigm Function, but allow the expression of facts about the language's inflectional morphology that go beyond the more straightforward realisation rules.

The realisation rules are arranged into blocks, with the rules in each block competing to determine which rule will apply next in the sequence of rule application that determines the word's inflected form (Stump 2001, 44). In affixal morphology, for example, one rule block typically determines the form of one affix 'slot'. The blocks are therefore ordered, but the rules within them are not. Pānini's principle is used to determine which rule of each block applies: the narrowest applicable rule (the one that applies to the smallest class of words, or that specifies most morphosyntactic properties) will apply. This is sometimes referred to as the principle of narrowness. The Identity Function Default (IFD) is a member of all rule blocks, and states that the default form is identical to the input: it is always the least narrow rule in every rule block. The IFD and Pānini's principle are claimed to operate universally, across all languages.

A simple example of PFM rules for the future and past tenses of Swahili verbs is given here, simplified from Stewart and Stump (2007, 402).

Block A	a.	$X_V\sigma$:{TNS:fut}	→ *ta*X
	b.	$X_V\sigma$:{TNS:past}	→ *li*X
	c.	$X_V\sigma$:{POL:neg, TNS:past}	→ *ku*X
Block B	d.	$X_V\sigma$:{AGR(su):{PER:1, NUM:sg} }	→ *ni*X
	e.	$X_V\sigma$:{AGR(su):{PER:2, NUM:sg} }	→ *u*X
	f.	$X_V\sigma$:{AGR(su):{PER:3, NUM:sg, GEN:{1,2}} }	→ *a*X
	g.	$X_V\sigma$:{AGR(su):{PER:1, NUM:pl} }	→ *tu*X
	h.	$X_V\sigma$:{AGR(su):{PER:2, NUM:pl} }	→ *m*X
	i.	$X_V\sigma$:{AGR(su):{PER:3, NUM:pl, GEN:{1,2}} }	→ *wa*X
Block C	j.	$X_V\sigma$:{POL:neg}	→ *ha*X

The Identity Function Default also acts as the least narrow member of any rule block:

IFD:	$X_V\sigma$:{}	→ X

To work out the correct form of the verb, we work through the rule blocks, applying the narrowest applicable rule according to Pānini's principle. For example, let us derive the first person future of the verb *taka* 'want', with positive polarity. From Block A, we apply rule (a) to the stem *taka* to get *ta-taka*. From Block B, we apply rule d, to get *ni-ta-taka*. In Block C, the Identity Function Default applies, giving the final form, *ni-ta-taka*.

Although a rule block in a given language may introduce affixes associated with different features, generalisations can often be made: for example, in rules for verbs, one rule block may contain rules that realise tenses, while another may contain rules that realise person and number combinations. This reflects tendencies in the structure of the language's morphology, without claiming that such tendencies are universal or exceptionless.

I now turn to a consideration of PFM's general theoretical advantages, and the reasons for implementing it in this work.

2.2.2 Advantages of PFM and justification for its use

PFM is a flexible and powerful model of morphology that can capture many kinds of morphological systems. Most obviously, it is suitable for languages with complex systems of inflection which go beyond simple concatenation: this includes languages with stem alternations, in which a single lexical item has multiple stems, which may be associated with different morphological contexts. In PFM, the correct stem is selected by the realisation rules. A simple example of stem alternations is found in Bulgarian: verbs typically have one stem used in the present, imperfect, imperative and gerund, and one used in the aorist, past passive participle, and verbal substantive (Stump 2001, 34), which may differ in accentuation, as in the verb *krad* 'steal', which has the stems *krad* and *krád* (Stump 2001, 39). There may also be much greater distinctions between stems, particularly in irregular lexemes, such as the Latin verb *ferre* 'carry' which has the stems *fer*, *tul* and *lāt* (Finkel 2010, 14).

Implementing PFM allows simple and effective analyses of autonomous morphological phenomena, such as morphomes, which can be defined as 'functions, lacking any inherent connection with a specific form or a specific meaning, which nonetheless serve systematically to relate form and meaning' (Maiden 2016b, 708). A typical example here is the distribution of continuants of the Latin perfective forms as stems across the verb paradigms of many Romance languages, including in Spanish, where this distribution, which is no longer transparently motivated by any morphosyntactic property, is sometimes called PYTA (for *pretérito y tiempos afines*, 'perfect and related tenses') (Maiden 2011, 174–201). The arbitrary nature of some morphological rules can also be straightforwardly accounted for in PFM by the recognition of inflection classes that capture shared patterns of inflection that are otherwise unmotivated. In Taylor's (2012) analysis of the verb

morphology of Spanish, for example, she recognises an inflection class called PYTA which includes verbs demonstrating stem alternations in this pattern.

As well as providing a straightforward analysis of complex inflectional morphologies, including both concatenative and nonconcatenative systems, PFM can equally be applied to languages with little inflectional morphology, such as English: some brief examples of applying PFM principles to aspects of English are given by Bonami and Stump (2016). Being able to capture morphological systems of very different types makes it easy to compare these systems, clarifying their similarities or differences. As well as the obvious benefits in typological or cross-linguistic work, this is a useful feature when considering regional variation or diachronic change in the morphological systems of a single language, since a side-by-side PFM comparison of rules makes it clear which rules have been extended, deleted or changed from one variety to another. I make use of this property of PFM particularly in Chapter 5, which considers the diachronic development of the double past participle phenomenon in European and Brazilian Portuguese.

Unlike many other theories of morphology, the concept of a 'morpheme' is irrelevant to PFM: there is no conception of a form-meaning pair below the level of the word, as only fully inflected forms are associated with morphosyntactic property sets. This aligns with the Lexical Integrity Principle of LFG, by which terminal nodes must correspond to fully inflected words, rather than to morphemes or other sub-word elements.

The power and flexibility of the PFM system has also been extended in various directions. Some of these extensions are discussed at greater length in section 3.3 below. One particularly interesting development is the extension of the model to include derivation and other types of lexical relatedness, in Spencer's (2013) theory of Generalized Paradigm Function Morphology (GPFM). Other developments include Stump's (2015) PFM2 model, which develops the insights of PFM to include disruptions in the mapping between syntax and morphology. Since this mapping is treated in Dalrymple's (2015) existing model of the interface between syntax and morphology in LFG, I

do not attempt to implement PFM2 in relation to LFG. On the other hand, I do implement some adaptations of the PFM model developed by Taylor (2012), which extend the standard PFM treatment of stem formation and stem selection.

A final reason for using PFM as the morphological theory for this thesis is its formal theoretical compatibility with LFG. As described in section 2.3, Dalrymple's (2015) description of the interface of morphology with syntax in LFG, including the treatment of the lexicon, is consistent with a realisational view of morphology. PFM also shares with LFG a commitment to being formally explicit and rigorously testable, as well as computationally implementable. In the next section, I discuss the integration of LFG and PFM used in this thesis.

2.3 Combining LFG and PFM

So far in this chapter, we have seen separately the benefits of the LFG architecture for studying syntax and other areas of the grammar, and the advantages of the PFM approach to morphology. In addition, we have seen that the two theories share theoretical principles, which favours a unified approach combining the two theories. In this section, I argue that combining the two theories offers an opportunity to integrate effective PFM analyses of autonomously morphological phenomena, as well as detailed accounts of other aspects of inflectional morphology, into the LFG architecture. This is important because the LFG architecture aims to make possible a full account of grammatical phenomena at all levels of structure, but in most implementations, a detailed theory of morphology is lacking. Overall, combining LFG with PFM enables a wider-reaching and more comprehensive account of the grammar than using LFG or PFM alone. To this end, in this section I set out some principles of the interface between syntax and morphology used in this thesis, following Dalrymple (2015).

In the version of LFG that I assume here, the syntax-morphology interface is not a function from one component to another (unlike, for example, φ, the function that relates c-structure nodes to their corresponding f-structure). The interface is instead made up of various functions and

relations between different types of information, which together describe mutual constraints between the syntax and the morphology. This model is described in detail by Dalrymple (2015); here I follow the up-to-date presentation set out by Dalrymple, Lowe and Mycock (2019, chap. 12).

In this section, I give an overview of the different parts of the system of constraints required between f-descriptions and morphological features. Where relevant, I give examples from the analysis of a straightforward Portuguese sentence (8):

(8) O João levava o bolo.
the.MSG João carry.IPFV.3SG the.MSG cake.MSG
'João was carrying the cake.'

The overall aim of the syntax-morphology interface is to produce the correct set of lexical entries for the language, referred to as $\mathscr{L}$. Lexical entries link many different pieces of information, as illustrated in Figure 3 (Dalrymple, Lowe, and Mycock 2019, 444).

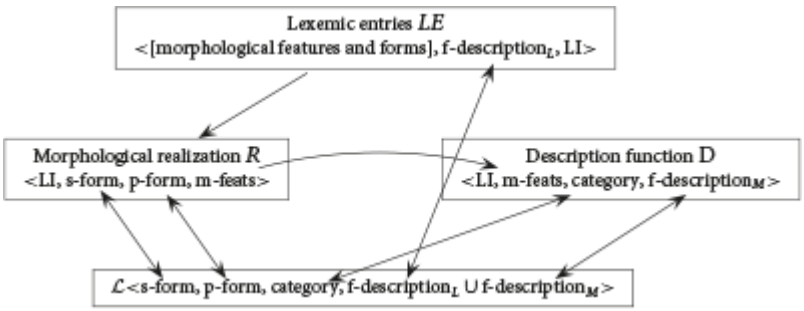


Figure 3 The set of lexical entries $\mathscr{L}$ and its requirements

In the following sub-sections I define and illustrate the features, functions, concepts and properties that appear in Figure 3 and are relevant to the interface between syntax and morphology.

2.3.1 Lexemic entries

The lexemic entry (LE) 'contains information of the kind which we would expect to find in an ordinary dictionary' (Dalrymple, Lowe, and Mycock 2019, 439). It takes the form of a three-place relation including the following pieces of information (Dalrymple, Lowe, and Mycock 2019, 440), which I exemplify with the Portuguese verb *levar* 'carry':

i. the form of the root and any non-predictable stem forms, together with morphological features encoding information about inflectional class, category, and other unpredictable features;
   [for example: {M-ROOT:lev, M-ICLASS:1, M-CAT:VERB}]
ii. an f-description that encodes syntactic, semantic, and other information associated with all wordforms of the lexeme, filling the role of Spencer's SYN and SEM[5];
   [for example: {(↑PRED FN)=TAKE}]
iii. the Lexemic Index, which we abbreviate as LI.
   [for example: LEVAR1]

The set in (i) can be characterised as a set of morphological features, including several types of information: the form of the root and the form of non-predictable stems, inflection class membership, and morphological category.

The f-description in (ii) includes the attribute PRED. Typically the value of PRED is the semantic form of the lexeme (i.e. the meaning associated with it), which for clarity I give in English even when dealing with data from other languages. In the case of verbs, the PRED value contributed also contains an argument list, given in angle brackets (e.g. {(↑PRED)='TAKE<SUBJ, OBJ>'}. The attribute FN can also be used to refer to the 'function component' of the semantic form (e.g.

[5] This refers to concepts from Spencer's (2013) *Lexical Relatedness*, discussed below in section 3.3.2.

{(↑PRED FN)='TAKE'}), with no argument list (Dalrymple, Lowe, and Mycock 2019, 154). This is used in the lexemic entries of verbs, where the same lexemic entry is associated with different possible argument lists: for example, the active {(↑PRED)='TAKE<SUBJ, OBJ>'} and the passive {(↑PRED)='TAKE<OBL_{AGENT}, SUBJ>'} (Dalrymple, Lowe, and Mycock 2019, 441). Referring to PRED FN in the lexemic entry for verbs allows correct information to be given for all forms of the lexeme, while the lexical entry for an individual wordform would refer to PRED with the relevant argument list.

The f-description may include further types of information, such as information structure, as well as the expected syntactic and semantic features. Any features encoded here must be applicable to *all* wordforms of the lexeme. They are thus distinct from the m-features that are part of the m-entry (see section 2.3.3). For the time being, I omit semantic information from lexemic entries. The representation of this information is discussed in section 3.2.1.

Finally, the Lexemic Index (iii) is defined as a 'unique identifier' associated with a lexemic root. In Dalrymple, Lowe and Mycock's (2019) presentation, mnemonic LIs such as DOG1 for *dog/dogs* are used. The LI is distinct from the semantic form of the lexeme, although when dealing with English it may look similar. Because the LI is only visible to the morphological component, not contributing any semantic or phonological information, what LI to use is essentially an arbitrary decision: an LI could equally well be just a number. Here I use an LI that reflects the surface form (rather than a translation), with a number. In PFM, the lexeme, L, in the pairing <L, σ> is identified in a similar way, (e.g. DOG), or in ambiguous cases, with a number, as in Stump's (2016 p93) example $RING_1$ 'put or make a ring around', past tense *ringed*, and $RING_2$ 'to (cause to) sound', past tense *rang*.

The general form of the lexemic entry in LFG is therefore:

(9) LE <{morphological information i.e. root, stems, category, other features}, {(f-description)}, Lexemic Index>

The standard lexemic entries (LE) that are associated with each element in our example sentence (8) are as follows:

(10) João: LE <{M-ROOT:JOÃO, M-ICLASS:1, M-CAT:NOUN}, {(↑PRED)='JOÃO'}, JOÃO1>

(11) levava: LE <{M-ROOT:LEV, M-ICLASS:1, M-CAT:VERB}, {(↑PRED FN)=TAKE}, LEVAR1>

(12) bolo: LE <{M-ROOT:BOLO, M-ICLASS:1, M-CAT:NOUN}, {(↑PRED)='CAKE'}, BOLO1>

(13) o: LE<{M-ROOT:O, M-ICLASS:1, M-CAT:DET}, {(↑DEF)=+}, O1>

Here I use arbitrary inflection class indices and simplified morphological root information, for indicative purposes. Any suggested changes to the treatment of stems or inflection class information, or other similar morphological properties, will involve additions or changes in these relations. Dalrymple, Lowe and Mycock note that (2019, 441–42):

> 'Depending on the theory of morphology that is assumed, different kinds of morphological information specified by different sets of m-features may be required. Again, since our analysis is compatible with a range of realisational morphological theories, we are agnostic about the content and representation of morphomic features such as inflectional class.'

In my proposals integrating PFM with LFG, a range of purely morphological information is listed in the lexemic entry, including inflection class membership, unpredictable stems, and some theme vowels. This is discussed and exemplified in greater detail for Portuguese verbs in Chapter 4.

2.3.2 Lexical entries

The lexical entry $\mathscr{L}$ is a relation between an inflected wordform and other pieces of information. As stated above, the purpose of the morphological component is to produce a complete set of fully specified and correct lexical entries.

Following Dalrymple and Mycock (2011), this takes the form of a four-piece relation:

(14) $\mathscr{L}$<s-form, p-form, category, f-description>

The s-form is 'the lexical representation specifying the syntactic string associated with a word' and the p-form 'the lexical representation specifying the prosodic string contributed by that word' (Dalrymple, Lowe, and Mycock 2019, 437). Category here refers to the word's c-structure category. For my purposes, the p-form is not relevant, so I generally omit the p-form, giving a three-place relation:

(15) $\mathscr{L}$<s-form, category, f-description>

The f-description information included in the lexical entry for a wordform ($\text{f-description}_L \cup \text{f-description}_M$) is distinct from that included in the lexemic entry (f-description_L). The f-description_L is true for every wordform of the lexeme (e.g. PRED value), but the f-description in the lexical entry is all the relevant information pertaining to the individual wordform. The f-description_M includes values for grammatical features such as, for verbs, SUBJ PERSON, SUBJ NUMBER, TENSE/ASPECT/MOOD, etc. and for nouns, NUMBER, GENDER and CASE, all of which are realised through interpretation of m-features.

For the sake of clarity, the f-description is abbreviated in the lexical entry as being of the form (↑PRED), rather than the technically more accurate $\varphi(\pi\ (\bullet))$[6].

[6] As shown in the architecture above, the function π maps information from s-string to c-structure; φ is the function from c-structure to f-structure.

The symbol ↑ is also used in LFG in annotations on c-structure trees (phrase structure annotations). The meaning demonstrated here is different from in that use.

Information about inflection class (relating to the morphological stem) is not included in the lexical entry, because it is not syntactically relevant: it is purely morphological or 'morphomic' (Dalrymple, Lowe, and Mycock 2019, 440). For the purposes of the lexical entry, it is irrelevant how the wordform gets its morphological shape.

Sample lexical entries:[7]

(16) João: $\mathscr{L}$<João, N, {(↑PRED)='JOÃO', (↑GEND)=MASC}>

(17) levava: $\mathscr{L}$<levava, V, {(↑PRED)='TAKE<SUBJ, OBJ>', (↑SUBJ NUM)=SG, (↑SUBJ PERS)={1|3}, (↑TNS = IPFV) (MD = IND)}>

(18) bolo: $\mathscr{L}$<bolo, N, {(↑PRED)='CAKE', (↑GEND)=MASC, (↑NUM)=SG}>

(19) o: $\mathscr{L}$< o, Det, {(↑DEF)=+, (↑GEND)=MASC, (↑NUM)=SG }>

To make more explicit the similarities and differences between the lexemic entry and the lexical entry, let us consider the lexemic entry for the verb *levar* 'carry' and the lexical entry for the inflected verb form *levava* 'was carrying', repeated together here for clarity.

(20) LE <{M-ROOT:LEV, M-ICLASS:1, M-CAT:VERB},
{(↑PRED FN)=TAKE},
LEVAR1>

(21) $\mathscr{L}$ <levava,
V,
{(↑PRED)='TAKE<SUBJ,OBJ>',
(↑SUBJ NUM)=SG, (↑SUBJ PERS)={1|3}, (↑TNS = IPFV) (MD = IND)}>

[7] The feature IPFV (imperfect) is an abbreviation of the Tense/Aspect/Mood value 'imperfect indicative'.

As we can see, the lexemic entry contains information about the morphology relevant to the whole lexeme, including the identity of the morphological root, inflection class membership and the morphological category specification 'verb', as well as an f-description containing the PRED FN value associated with the verb, and the Lexical Index. By contrast, the lexical entry of the inflected form contains the syntactic category V and a much more complete f-description containing all the relevant grammatical features of the inflected form, as well as the semantic form (PRED value with relevant argument list).

Now we have defined and described some of the important pieces of information associated with words from the LFG perspective, we must consider how these are related to each other. I start by considering the relation $\mathscr{R}$.

2.3.3 M-entries and the morphological realisation relation $\mathscr{R}$

The morphological realisation relation $\mathscr{R}$ is a vital part of the syntax-morphology interface. It encodes the relation between a wordform, its meaning and its associated morphological features: '$\mathscr{R}$ is simply a means of associating m-features with p-forms and s-forms relative to a lexemic root' (Dalrymple, Lowe, and Mycock 2019, 442). The $\mathscr{R}$ relation for a language is defined as a set of morphological entries (m-entries): four-place relations that associate a Lexemic Index with its s-form, p-form and m-features. In my analysis, the internal structure and definition of $\mathscr{R}$ is given by PFM (as suggested by Dalrymple, Lowe, and Mycock 2019, 442–43, 456), which provides the means of linking the LI to the s-form, p-form and m-features. Beyond the suggestion to implement PFM, it is often stated that LFG is compatible with various approaches to morphology, including 'realisational, constructional or morpheme-based' approaches (Sadler and Nordlinger 2018, 222), but there are few, if any, fully fleshed-out models. As mentioned in section 1.2, LFG morphology

work tends to focus on the interface between syntax and morphology, rather than 'morphological theory proper' (Sadler and Nordlinger 2018, 214).[8]

We can define $\mathscr{R}$ for each terminal node on the tree to give m-entries for each word. The general form of an m-entry is:

(22) $\mathscr{R}$ <LI, s-form, p-form, m-features>

Again, though, the phonological form is irrelevant to my concerns, so I omit it:

(23) $\mathscr{R}$ <LI, s-form, m-features>

The m-entries I propose for our sample sentence are therefore:

(24) João: $\mathscr{R}$ < JOÃO1, João, {M-CAT:NOUN, M-GEND: MASC, M-NUM:SG} >

(25) levava: $\mathscr{R}$ < LEVAR1, levava, {M-CAT:VERB, M-TENSE:IPFV, M-PERS: 3, M-NUM: SG, M-VOICE: ACTIVE}

(26) bolo: $\mathscr{R}$ <BOLO1, bolo, {M-CAT:NOUN, M-GEND: MASC, M-NUM: SG}

(27) o: $\mathscr{R}$ <O1, o, {M-CAT:DET, M-GEND: MASC, M-NUM: SG}

Portuguese nouns do not inflect for case, so there is no feature M-CASE in the m-entries for the nouns *João* and *bolo*. It would be unjustified to claim that *João* is morphologically nominative or subjective and *bolo* is morphologically accusative or objective. The syntactic case properties of nouns and pronouns in Portuguese are briefly discussed in section 3.1.2 below.

[8] As mentioned above, an example of recent work based on LFG using a different (Distributed Morphology) approach is Melchin, Asudeh and Siddiqi (2020).

M-features and their relationships to corresponding syntactic features are discussed further in the next section, which looks at the description function $\mathcal{D}$.

2.3.4 The description function $\mathcal{D}$

The next part of the syntax-morphology interface that I consider is the description function $\mathcal{D}$. This 'maps a set of m-features provided by $\mathscr{R}$ to the appropriate c-structure category and f-description$_M$ for a word form' (Dalrymple, Lowe, and Mycock 2019, 443). In more detail:

> 'The description function $\mathcal{D}$ constitutes the interface between the morphological component and the rest of the grammar, since it defines the relation between the m-features for a word form, which are invisible outside the morphological module, and the phrase structure category and f-description that appear in its lexical entry.' (Dalrymple, Lowe, and Mycock 2019, 457)

The general form of the description function $\mathcal{D}$ is (Dalrymple, Lowe, and Mycock 2019, 443):

(28) $\mathcal{D}$ <Lexemic Index, m-features, category, f-description$_M$ >

More specific examples for our sentence are:[9]

(29) João: $\mathcal{D}$ < JOÃO1
{M-CAT: NOUN, M-GEND: MASC, M-NUM:SG},
N,
{(↑NUM)=SG, (↑GEND)=MASC}>

[9] A full analysis should locate grammatical features in the feature bundles INDEX and CONCORD in the f-description, but for simplicity of presentation, this information is omitted here. A full discussion of INDEX and CONCORD is given in Chapter 6.

(30) levava: $\mathcal{D}$ < LEVAR1

{M-CAT: VERB, M-TENSE: PAST 1, M-PERS: 3, M-NUM: SG, M-VOICE: ACTIVE}[10]

V,

{(↑SUBJ NUM)=SG, (↑SUBJ PERS)= {1|3}, (↑TENSE)= PAST, (↑ASPECT)= IMPF, (↑VOICE)= ACTIVE}>

(31) bolo: $\mathcal{D}$ < BOLO1

{M-CAT: NOUN, M-GEND: MASC, M-NUM:SG},

N,

{(↑NUM)=SG, (↑GEND)=MASC}>

(32) o: $\mathcal{D}$ < O1

{M-CAT: DET, M-GEND: MASC, M-NUM: SG}

D,

{(↑DEF)=+}>

These examples clarify the purpose of $\mathcal{D}$: to describe the direct relationship between the morphological features (such as M-CAT, M-GEND) and their syntactic correlates (such as category, f-structure gender feature, etc.), which can either be straightforward, or more complex.

[10] In this example, I have given separate specifications of TENSE and ASPECT in the syntax, but followed Taylor (2012) in combining these two features in the morphology (see section 3.3.1). Under this analysis, the morphological tense/aspect combination represented here has the feature M-TENSE: PAST 1.

According to Dalrymple, Lowe and Mycock (2019, 443), $\mathcal{D}$ corresponds with the 'rules of paradigm linkage' in the PFM2 theory (see section 3.3.3). This overlap is one of the reasons why I implement the original PFM rather than PFM2 in the context of the LFG architecture.

2.3.5 The relationship between f-features and m-features

As mentioned above, the relationship between morphological features and their syntactic correlates is not always straightforward. In the default situation, for example, there is a match between the m-feature M-GEND and the GEND feature at f-structure. But in other cases a mismatch occurs, meaning there is no simple mapping between the m-features and syntactic features.

Not every m-feature has a correlate at the level of f-structure. For example, inflection class, represented as M-ICLASS, is not relevant outside the morphological component, and does not have an f-structure correlate. The category information typically represented as M-CAT at the level of m-structure is also not explicitly represented at f-structure.

However, some features represented at f-structure as properties of the arguments of the main predicate are relevant to the morphology of that predicate. In our example sentence, the verb *levava* has third or first person singular marking, meaning that we know from its morphological form that the subject is third person and singular or first person and singular. As we saw above, this information is captured in LFG by the description function $\mathcal{D}$, which maps the relevant part of the lexical entry (e.g. for person, {(↑SUBJ PERS) = 1) | (↑SUBJ PERS) = 3)}) to the f-description$_M$.[11]

The term 'm-features' can be applied to two different sets of information. In the lexemic entry, m-features include lexeme-general, purely morphological, 'morphomic' features, such as inflection

[11] The f-description$_M$ is the morphologically encoded portion of a word's f-description, which contains information represented at the level of f-structure (see section 2.3.2).

class and morphological roots and stems. In the m-entry, m-features are specific to the inflected wordform, which have corresponding syntactic or semantic features, and are called morphosyntactic m-features or morphosemantic m-features. The m-feature M-CAT ('morphological category') appears in both these sets.

For Portuguese, the m-features that are specific to the inflected wordform would include (for nouns) number and gender, and (for verbs) number, person, tense, aspect,[12] mood, and voice. On the other hand, m-features that are shared by the whole lexeme include inflection class (M-ICLASS) and information about the morphological roots/stems (M-ROOT and M-STEM). The LFG architecture thus distinguishes between types of m-features with different properties, especially in their implications outside the morphology. Such distinctions have been discussed at length in the LFG literature, going back to Sadler and Spencer (2001) and papers collected in Sadler and Spencer (2004), such as Frank and Zaenen (2004 [2002]) and Sells (2004).

As set out by Sadler and Spencer (2001), m-features and syntactic features (s-features) must be kept separate, even though there is frequently a clear mapping between them. They give the example of the French past tense *est venu* 'came':

(33) m-features: [m-Tense:Pres] + [m-Participle:Perfect][13]
s-features: [s-TENSE PAST]

[12] See section 3.3.1 for discussion of the possibility of representing tense and aspect together.

[13] Sadler and Spencer make a typographical distinction between m-features (represented in lower case) and s-features (in small capitals). When not quoting Sadler and Spencer, I follow Dalrymple, Lowe and Mycock (2019) in using small capitals for both.

Here, the morphological tense of the auxiliary verb *est* 'is' is reflected in the feature [m-Tense:Pres], while the syntactic tense value of the whole auxiliary plus participle construction is captured by the feature [s-TENSE PAST]. This is a typical example of a mismatch between morphological and syntactic features, which can be simply represented in the LFG architecture through principles of $\mathcal{D}$-mapping. These mismatches give a clear motivation for separating out morphological features from their syntactic and semantic correlates, allowing for either straightforward correspondences (with no mismatch), or for atypical mappings where we find mismatches. Building on previous LFG work on $\mathcal{D}$-mapping, including on 'affix hopping' and Latin deponent verbs, (Butt, Fortmann, and Rohrer 1996; Butt, Niño, and Segond 1996; Sadler and Spencer 2001; Frank and Zaenen 2002), I consider the relationship between morphological and syntactic features in the context of Portuguese verbal inflection in section 3.1.2, with a view to accounting for the distinctions between the morphological and syntactic tense features of past participles and the periphrases in which they are found.

2.3.6 Categories in LFG and PFM

LFG recognises several different lexical categories: nouns, prepositions, verbs, adjectives and adverbs.[14] As we saw above, lexical category can be treated as an m-feature in LFG. Like morphosyntactic m-features, it has a correlate in the syntax ('c-structure category' or 'syntactic category'), and there may be mismatches between the morphological category and the syntactic category: Dalrymple, Lowe and Mycock (2019, 451) quote Spencer's (2013) example of the Russian *stolovaja* 'dining room', which has the morphological (inflectional) behaviour of an adjective, but is a noun in its syntax.

[14] At the level of constituent structure, LFG also recognises 'functional' categories alongside lexical categories: I ('inflection'), C (complementisers) and D (determiners).

Unlike other m-features, M-CAT appears in both the lexemic entry and in the m-entry. Category is relevant for all forms of the lexeme, so intuitively belongs in the lexemic entry, but it also needs to be interpreted by the description function $\mathcal{D}$ to give the syntactic category, so must be in the m-entry. For example, a morphological category M-CAT:NOUN would typically be interpreted to give the syntactic category N. Unlike other m-features in the m-entry, though, M-CAT does not generally vary across different wordforms. However, this claim is not as simple as it initially seems. Stump (2016, 59) notes that:

> 'A lexeme's syntactic category membership determines that of the forms that realize it. This does not entail that all of the forms realizing a lexeme necessarily belong to the same syntactic category, since there are familiar instances in which that is not the case; for instance, the realizations of verbal lexemes can be plausibly assumed to include participles and infinitives, even though in many languages, these are respectively adjectival and nominal rather than verbal.'

This insight is treated in detail in Spencer's Generalized Paradigm Function Morphology (2013), capturing the many types of lexical relatedness beyond straightforward inflection, such as the relationship between verbs and their non-finite forms, using a formalism that builds on PFM (section 3.3.2).

In PFM, category information is not treated as a morphosyntactic feature like person/number/gender or tense/aspect/mood; neither is it treated in the same way as a morphomic feature like inflection class. In a standard PFM approach, the rules of exponence apply to lexemes from a particular category, and are annotated with the relevant category.

In some ways, PFM makes a different distinction from that made in LFG: category information is not an m-feature, but a different type of information altogether, which conditions what rules will apply. LFG also makes a distinction not made in PFM: the LFG conception of M-CAT distinguishes a word's inflectional (morphological) category from its syntactic category, and accounts for this

mismatch in the same way as other mismatches between morphology and syntax. However, it is possible to reconcile these different approaches in a unified LFG-PFM treatment of morphology and syntax, by incorporating aspects of both treatments of lexical category as required by the separate modules, as I show for Portuguese past participles in section 4.2.

Questions of lexical category are discussed at greater length in Chapter 4, on Portuguese double past participle forms, with a particular focus on the lexical category properties of past participles and their position in relation to verbal inflectional paradigms, as well as in Chapter 6, looking a the category properties of nouns versus pronouns.

2.4 Chapter summary

As we have seen in this chapter, implementing PFM within the LFG architecture provides a coherent, powerful and flexible theoretical approach to morphological and syntactic phenomena. The two approaches share many conceptual similarities regarding the nature of the grammar and the role of linguistic formalisms. In addition, Dalrymple (2015) provides a clear model of the interface between morphology and syntax, which sets out in a formally explicit way the relationship between these two parts of the grammar. PFM can be straightforwardly integrated into this model as the morphological realisation relation $\mathscr{R}$ to provide the fully inflected forms, paired with morphosyntactic property sets, that are needed to give the language's full set of lexical entries. This allows detailed accounts of purely morphological phenomena to be integrated into wider LFG approaches to the grammar. I refer to this combined approach, which integrates PFM with LFG, as LFG-PFM.

More specifically, in this thesis, applying the LFG-PFM model provides clear insights into the case studies of divergence and mismatches in Portuguese morphology and syntax. As briefly mentioned in section 2.1, and developed further in the next chapter, the LFG parallel architecture allows novel insights into grammatical change, because different aspects of change can be represented

separately. This allows the different factors to be distinguished and their relationships to be investigated.

In the next chapter, I look in more detail at how LFG and PFM can be applied to analyse Portuguese morphology and syntax, demonstrating some important properties of these theories and their extensions.

Chapter 3 Applying LFG and PFM to Portuguese

Having seen the basic elements of LFG and PFM, along with the interface between syntax and morphology used in the LFG architecture, this chapter now applies LFG-PFM insights to the analysis of Portuguese. The aims of the chapter are, first, to demonstrate some basic features of Portuguese syntax and morphology from an LFG-PFM perspective, and second, to look at some extensions or other aspects of LFG and PFM theory. In both cases I particularly focus on areas that are relevant to the treatment of mismatches. This chapter therefore sets the scene for the Portuguese case studies in Chapter 4, Chapter 5 and Chapter 6, in which the LFG-PFM theory is used to analyse morphosyntactic phenomena from the history of Portuguese.

The chapter begins by presenting some simple LFG analyses of Portuguese syntax (section 3.1.1), before considering the important features of Portuguese morphology and how these are represented in PFM (section 3.1.2). In section 3.2 I look at some further points of LFG theory: in section 3.2.1 I discuss the role of semantic structure in LFG, and in section 3.2.2 I describe applications of LFG to diachronic questions. Section 3.3 reviews four suggested extensions of PFM, some of which are implemented to a greater or lesser extent in the rest of the thesis. Finally, section 3.4 brings together conclusions from the chapter.

3.1 Analysis of Portuguese syntax and morphology

3.1.1 Basics of Portuguese syntax in LFG-PFM

This section gives partial LFG analyses of simple Portuguese phrases or sentences, to show the kind of analysis assumed in the rest of this thesis. The Romance languages have been well-studied from an LFG perspective, although there is relatively little LFG work on Portuguese in particular. Two exceptions are Luís's work on clitics (Luís 2004; Luís and Otoguro 2004; 2005) and de Alencar's computational implementation of LFG for a fragment of Brazilian Portuguese (2013).

Since Portuguese shares many syntactic properties and phenomena with other Romance languages, I do not propose to give a full LFG analysis of the syntax of Portuguese here.

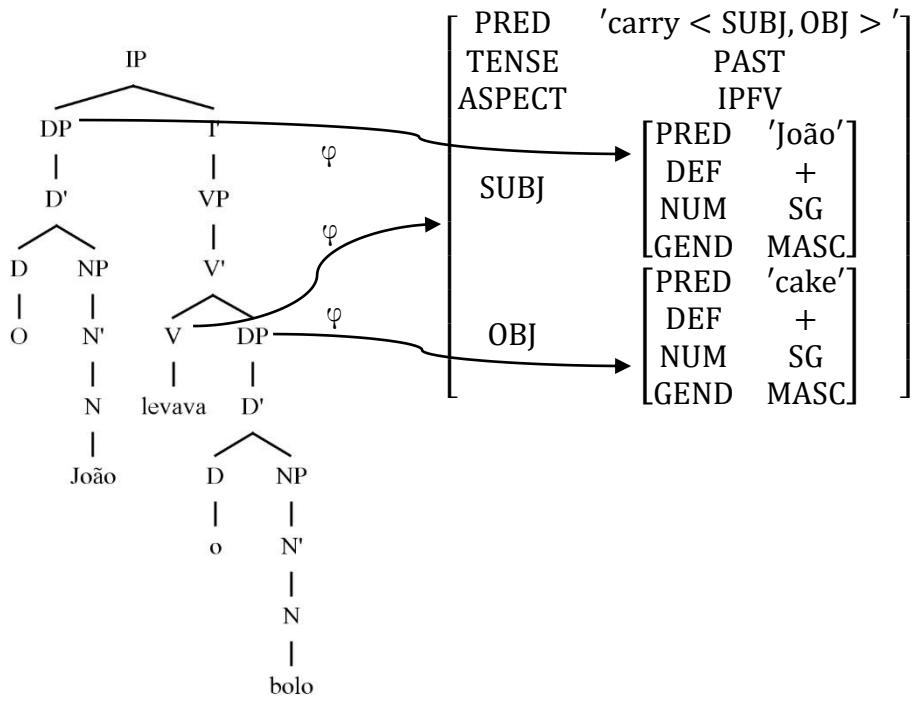
The focus of the thesis is on the interactions between syntax and morphology and their role in mismatch phenomena, so some aspects of the LFG architecture for syntax are more relevant than others. In particular, constituent structure does not play an important role in the rest of the thesis, since I am not concerned with issues of linear order or constituency. However, understanding the relationship between constituent structure and functional structure is important to understanding the LFG architecture more broadly. I present relevant information about constituent and functional structure in this section, including a discussion of two mismatch phenomena from Portuguese that raise questions in this respect. I also discuss other aspects of the LFG architecture in section 3.2 below.

In example (35), I present constituent structure and functional structure analyses of the simple sentence given in (34), which was also used for exemplification in Chapter 2. I also indicate the φ mapping between the two structures, which in this example is straightforward. Mismatches at this level of structure are not considered in the thesis, but Dalrymple, Lowe and Mycock (2019, chap. 4) give examples from several languages of possible mismatches in the mapping between constituent structure and functional structure, such as discontinuous constituents and clitic doubling.

In Portuguese c-structures, I follow the DP hypothesis, which is empirically motivated for Portuguese as proper nouns and possessives are frequently preceded by a determiner. I refer the reader to Dalrymple, Lowe and Mycock (2019, 100–101) for further information on the DP hypothesis.

(34)	O	João	levava	o	bolo
	the.MSG	João	carry.IPFV.3SG	the.MSG	cake.MSG
	‘João was carrying the cake’				

(35) C-structure and F-structure for *O João levava o bolo* 'João was carrying the cake'



In a straightforward example like this one, each terminal node of the constituent structure corresponds to one lexical item. However, this is not always the case. There can be a mismatch between the number of syntactic words in the syntactic string (s-string: a representation of syntactic words in their linear order), and the number of terminal nodes at the level of constituent structure. This can be seen clearly in the analysis of a simple Portuguese phrase including a contracted form of a preposition and the definite article:

(36) pelo menino

by.the.MSG boy.MSG

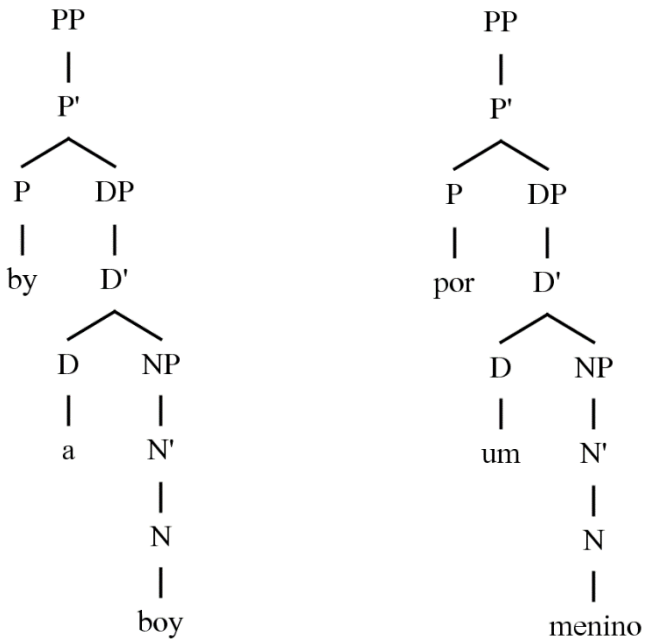
'by the boy'

In Portuguese, like in English, the constituent structure of a prepositional phrase includes a preposition, a determiner, and a noun phrase. This can clearly be seen in example (37), which has an indefinite article rather than the definite article, but is otherwise identical in meaning.

(37) por um menino
by a.MSG boy.MSG
'by a boy'

The c-structures of this phrase and its English equivalent are given in (38), and an f-structure in (39).

(38) C-structure representations of English prepositional phrase 'by a boy' and Portuguese prepositional phrase *por um menino* 'by a boy'



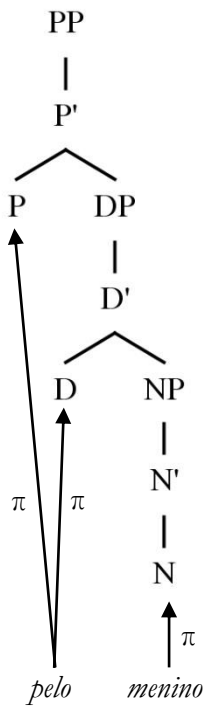
(39) F-structure representation of prepositional phrase *por um menino* 'by a boy'

$$\begin{bmatrix} \text{PRED} & \text{'by} < \text{OBJ} > \text{'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'boy'} \\ \text{DEF} & - \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{M} \end{bmatrix} \end{bmatrix}$$

This raises the question of how to capture the constituent structure of example (36), where both the P node and the D node correspond to the single lexical item *pelo* 'by.the.MSG'. This and similar

forms cannot be accounted for by regular phonological rules: the form *pelo* represents a contraction of *por* 'by' and *o* 'the'. This type of syntactic behaviour is described and analysed from an LFG perspective by Wescoat (2002; 2005), who distinguishes between the 'linear structure' and the constituent structure. In the most modern LFG terminology, the linear structure corresponds to the s-string. In Wescoat's theory of lexical sharing, a single item in the s-string can be associated with two terminal nodes in the c-structure. This is indicated in the constituent structure representation in (40). The arrows linking the annotations at the bottom of the constituent structure to the terminal nodes represent the π projection from the s-string, which encodes linear order, to the terminal nodes of c-structure, which encodes hierarchical relations. This use of arrows is a more precise and explicit representation of the π projection (compared with the use of simple lines as in previous c-structures), but for simplicity I do not use this representation in the rest of the thesis, instead annotating c-structures in the standard way.

(40) C-structure representation of *pelo menino* 'by the boy' with lexical sharing



by.the boy

(41) F-structure representation of prepositional phrase *pelo menino* 'by the boy'

$$\begin{bmatrix} \text{PRED} & \text{'by} < \text{OBJ} > \text{'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'boy'} \\ \text{DEF} & + \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{M} \end{bmatrix} \end{bmatrix}$$

LFG thus offers a simple account of the relationship where a single word in the string corresponds to two different nodes at the level of constituent structure.

Another mismatch phenomenon in the morphosyntax of Portuguese that presents a challenge to LFG is the clitic system. Portuguese clitics have been the subject of quite a significant amount of attention within LFG (Luís and Sadler 2003; Luís 2004; Luís and Otoguro 2004; 2005) as well as from other theoretical perspectives including HPSG (Crysmann 1997) and transformational generative approaches (Madeira 1992; I. Duarte and Matos 2000; Galves, Moraes, and Ribeiro 2005). Here I briefly discuss some aspects of these forms, focusing on LFG approaches. Accounting for European Portuguese clitics in a way that is consistent with the Lexical Integrity Hypothesis presents a particular challenge.

Romance clitics are generally assumed to be verbal affixes of some kind. In Portuguese, however, clitics can occur either before the verb (proclisis), inserted within the verb (mesoclisis) or after the verb (enclisis), with the same forms fulfilling the same grammatical function regardless of their order in relation to the verb.[15] The position of the clitic depends on the syntactic context of the

[15] Portuguese clitics undergo morphophonological processes which affect their morphological form. They may be combined into clusters (e.g. *me* 1SG.INDOBJ + *o* 3SG.M.DOBJ combine as *mo*). Morphophonological rules also apply in clitics' interactions with some verb forms. For example, in (42), the infinitive verb *descrever* 'describe' followed by the (enclitic) *o* 3SG.M.DOBJ combine to

verb. In general, enclisis is treated as the default position, with proclisis being triggered by particular contexts, as shown in (42) and (43). Mesoclisis (44) occurs in formal registers with the future or conditional tense. These examples are from the *Corpus do Português* (Web/Dialects).[16]

(42) Não tentarei descrevê -lo, até
NEG try.FUT.1SG describe.INF 3SG.DOBJ.MSG even
porque não o conseguiria fazer.
because NEG 3SG.DOBJ.MSG achieve.COND.1SG do.INF
'I won't try to describe it, because I wouldn't even manage to do it.'

(43) Os avós desta vez não lhe
the.MPL grandparents.PL of.this time NEG 3SG.INDOBJ
deram peúgas, mas deram -lhe
give.PAST.3SG sock.PL but give.PAST.3PL 3SG.INDOBJ
um pijama!
a.MSG pyjama
'This time his grandparents didn't give him socks, but gave him pyjamas!'

give the form *descrevê-lo*. For the sake of simplicity, I do not present the full details here, but refer the reader to Luís (2004) for further information.

[16] As described in the List of Abbreviations, in the glosses I follow standard Portuguese orthographical conventions in the use of hyphens to separate postverbal clitics from verbs.

(44)	cada	mês	os	nossos	leitores	ajudar
	each	month	the.MPL	our.MPL	readers.MPL	help.INF
	-nos		-ão	a	escolher	um	filme
	-1PL.DOBJ		FUT.3PL	to	choose.INF	a.MSG	film.MSG

'Every month our readers will help us to choose a film.'

In the first clause of (42), the negative marker *não* triggers proclisis, and in the first clause of (43), the subordinating conjunction *porque* 'because' triggers proclisis. There are many other kinds of triggers (Luís and Otoguro 2004, 336), which do not share any particular pragmatic, semantic or syntactic properties other than the fact that they trigger proclisis (Alsina Forthcoming). Additionally, when the clitic occurs before the verb, it can be separated from it by certain lexical items, particularly adverbs (Luís and Otoguro 2005, 267):

(45)	Eu	sei	que	ele	o
	1SG.SUBJ	know.PRES.1SG	that	3SG.SUBJ.M	3SG.DOBJ.M
	ainda	não	visitou		
	yet	NEG	visit.PAST.3SG		

'I know that he still has not visited him'

Naturally occurring examples of this construction can be found (*Corpus do Português*):

(46)	Também	eu	o	não	quis	divinizar
	also	1SG.SUBJ	3SG.DOBJ.M	NEG	want.PAST.1SG	deify.INF

'Also I didn't want to deify him.'

The pronoun *o* is clearly the same form as would be used in an enclitic position (immediately after the verb), in which case it would be treated as a straightforward verbal affix, formed in the

morphology as part of a single morphological unit with the verb.[17] This poses a problem, because verbal affixes should normally occur in only one linear order in relation to the verb (either as a suffix or as a prefix) and should not be separable from the verb by other lexical items. We therefore have a single morphological item (verb with affix) which corresponds to two c-structure nodes that are not adjacent, which conflicts with the Lexical Integrity Principle. This is similar in some respects to the mismatch between elements in the s-string and terminal nodes of the c-structure discussed above, but the question of nonadjacency adds further complexity.

Some adaptations of standard LFG assumptions are needed to allow a single morphological item (such as a combination of a verb stem and a clitic affix) to be mapped onto more than one syntactic node. One solution to analysing European Portuguese proclitics as phrasal affixes within LFG incorporates Wescoat's (2002) insights into the relationship between constituent structure and the string, taking the view that 'lexical integrity' refers to the morphological integrity of inflected forms, rather than being 'a condition on the mapping between words and c-structure terminals' (Luís and Otoguro 2005, 268). This allows a full analysis of the properties of the proclitic forms, capturing their constituent structure, functional structure and morphological properties.

---

[17] Here I follow Luís (2004, 345), who analyses the Portuguese forms usually termed 'enclitics' as affixes, stating that 'robust signs of morphological construction between the enclitic and the verb indicated to us that enclitics to a verbal stem [are] inflectional suffixes. The term 'clitic' therefore has been used merely as a descriptive cover term, not as a theoretical concept.' Luís recognises an asymmetry between enclitics, which are 'genuine verbal suffixes', and proclitics, which are 'phrasal prefixes'. Similarly, Luís and Spencer (2012, 146–47) argue that 'in terms of canonical properties', Romance pronominal clitics 'are rather closer to affixes than clitics'.

Having seen the basic elements of an LFG analysis of Portuguese syntax, in the next section I describe the fundamental parts of a PFM analysis of Portuguese morphology.

3.1.2 Basics of Portuguese morphology in LFG-PFM

As noted in Chapter 2, PFM is well-suited to analysing languages which have a rich inflectional morphology. There have been two previous PFM analyses of Portuguese. Luís (2004), as discussed in section 3.1.1 above, focuses on accounting for the behaviour of clitics, along with a small number of verbal tenses: the imperfect, future and conditional. Taylor (2012) presents a more wide-reaching analysis of Portuguese, as part of a comparative PFM analysis of six Romance languages: Catalan, French, Italian, Portuguese, Romanian and Spanish. Since Taylor's account includes a more comprehensive coverage of Portuguese verbal inflection, in this thesis I build on her account. In particular, in Chapter 4, I extend Taylor's PFM analysis to account for the double past participle phenomenon in Portuguese. Further details of Taylor's analysis, with a focus on her treatment of stems, are given in section 3.3.1.

To first provide some background on Portuguese verbs, I present their morphosyntactic properties here. The relevant types of inflection are summarised below.

- Finiteness: verbs may be finite, or non-finite (infinitive, past participle, gerund)
- Person and number: finite verbs inflect for person (1, 2 or 3) and number (singular or plural)[18]
- Mood: finite verbs may be indicative or subjunctive

[18] In the case of the personal infinitive, a non-finite form may also inflect for person and number. See section 4.4.3 for a brief discussion.

- Tense and aspect: finite verbs may be present, imperfect, future, past (perfect), pluperfect or conditional

For simplicity, I treat tense and aspect together, distinguishing the imperfect and past (perfect) forms as separate tenses, rather than treating them as realising an aspectual opposition. In this decision I follow Taylor's (2012) analysis of Romance generally and Portuguese specifically. This gives the set of tense and mood features for Portuguese verbs set out in Table 4.

Indicative (MOOD:IND)	**Subjunctive (MOOD:SBJV)**
Present *come* 'he eats'	Present *coma* 'he eats'
TNS:PRES	TNS:PRES
Imperfect *comia* 'he used to eat'/ 'he was eating'	Imperfect *comesse* 'he ate' (rare)
TNS:PAST 1	TNS:PAST 1
Future *comerá* 'he will eat'	Future *comer* 'he will eat'
TNS:FUT	TNS:FUT
Past *comeu* 'he ate'	-
TNS:PAST 2	
Pluperfect *comera* 'he had eaten'	-
TNS:PLUP	
Conditional *comeria* 'he would eat'	-
TNS:COND	

Table 4 Tenses and moods of Portuguese verbs

All of these forms are *simple* (simplex or one-word).[19] For the simplex forms, there is generally a straightforward correspondence between the morphological and the syntactic tense/aspect features. That is to say, the morphological feature 'TNS:PRES' corresponds to the syntactic feature (which we would see at the level of f-structure) 'TENSE' with the value 'PRES'. Similarly, the morphological feature 'MOOD:IND' corresponds to the syntactic feature 'MOOD' with the value 'IND'.[20] In the LFG syntax-morphology interface used here, these correspondences are captured in $\mathcal{D}$-mapping:

(47) $\overset{\mathcal{D}}{\text{M-TENSE:PRES}} \Rightarrow (\uparrow\text{TENSE}) = \text{PRES}$

(48) $\overset{\mathcal{D}}{\text{M-MOOD:IND}} \Rightarrow (\uparrow\text{MOOD}) = \text{IND}$

Since Portuguese, like many other languages, has more tenses than English, a more extensive range of syntactic tenses is required than the three proposed for English in LFG (PRES, PAST, FUT). Depending on the syntactic analysis of Portuguese tenses assumed, we could also recognise a more complex $\mathcal{D}$-mapping from the morphological tense PAST 1 to the syntactic features TENSE: PAST and ASPECT: IMPF, and from PAST 2 to TENSE: PAST and ASPECT: PERF. A breakdown of the features

[19] An alternative analysis holds that the future and conditional in European Portuguese are 'a kind of compound forms' (Dubert and Galves 2016, 425) since clitic pronouns can be positioned within the verb in mesoclisis (3.1.1). This is not compatible with a morphological approach to the Portuguese clitics, such as that provided by Luís (2004).

[20] These features, particularly mood, are often not included in f-structure representations, following the LFG convention to omit features that are not relevant to the discussion. However, for the sake of completeness, we should have an account of tense and mood at f-structure.

of the conditional and the pluperfect is also possible, although developing this idea fully is beyond the scope of my analysis.

Portuguese also has many possible compound or periphrastic verbal forms, some of which are more common than others. One analysis of Portuguese compound forms and periphrases includes the following (Dubert and Galves 2016, 426).

Forms with auxiliary *ser* 'be' + past participle: passive voice

(49) foi comido
be.PAST.3SG eat.PTCP.MSG
it was eaten

Forms with auxiliary *ter* 'have' + past participle: iterative aspect with present tense auxiliary, otherwise no particular aspectual meaning

(50) tenho comido
have.PRES.1SG eat.PTCP
'I have been eating'

(51) tinha comido
have.IPFV.1SG eat.PTCP
'I had eaten'

Forms with *estar* 'be' + preposition *a* 'to' + infinitive (in standard and northern European Portuguese): progressive aspect

(52) estou a comer
be.PRES.1SG to eat.INF
'I am eating'

Forms with *estar* 'be' + gerund (in Brazilian Portuguese and southern European Portuguese): progressive aspect

(53) estou comendo
be.PRES.1SG eat.GERUND
'I am eating'

Forms with *ir* 'go' + infinitive: future tense

(54) vou comer
go.PRES.1SG eat
'I am going to eat'

Forms with *ter* 'have' + preposition *de* 'of' or *que* 'that': obligation (mood)

(55) tenho de comer
have.PRES.1SG of eat.INF
'I have to eat'

Another analysis suggests that only forms with *ter* + past participle should be considered compound tenses, distinguished from other periphrastic forms because they can alternate directly with the tense of simple forms (Oliveira 2013, 528). The question of which forms should be counted as 'compound tenses' and which as 'periphrases' is not crucial to my discussion of any of these forms. A terminological distinction is sometimes made between forms which are part of the inflectional paradigm of a verb, and forms involving more than one lexical item which are created in the syntax. As discussed in section 3.3.4, I treat both structures with *ser* 'be' and a past participle and those with *ter* 'have' and a past participle as being members of the verbal paradigm, due to the clear parallelism between them.

A fuller range of periphrases with one of the auxiliary verbs *ser* 'be' and *ter* 'have' plus the past participle is demonstrated in Table 5 below, since these are considered in detail in Chapter 4 and Chapter 5.

Periphrases with *ter* 'have'	**Periphrases with *ser* 'be'**	**Tense of auxiliary**
iterative past *tem comido*	present passive *é comido*	present indicative
'he has been eating'	'it is eaten'	
compound pluperfect *tinha comido*	imperfect passive *era comido*	imperfect indicative
'he had eaten'	'it was eaten' (state)	
future perfect *terá comido*	future passive *será comido*	future indicative
'he will have eaten'	'it will be eaten'	
---	past passive *foi comido*	past (indicative)
	'it was eaten' (punctual)	
past conditional *teria comido*	conditional passive *seria comido*	conditional (indicative)
'he would have eaten'	'it would be eaten'	

Table 5 Periphrastic forms with *ter* 'have' and *ser* 'be'

The relationships between morphological and syntactic features in the periphrastic forms are clearly more complex than for the simplex forms shown above, necessitating more complex $\mathcal{D}$-mapping. As discussed in section 2.3.5, periphrastic forms are one of the key examples where m-features correspond to different s-features. Here I repeat Sadler and Spencer's analysis of the French past tense *est venu* 'came' (given as example (33) above):

(56) m-features: [m-Tense:Pres] + [m-Participle:Perfect]

s-features: [s-TENSE PAST]

In this analysis of the periphrasis, the morphological features of the present tense verb *est* 'to be' and the perfect participle *venu* 'came' combine to give a single syntactic tense feature 'PAST'. There

are several different proposals in LFG for how this $\mathcal{D}$-mapping takes place, often referring to the module of m-structure (Butt, Niño, and Segond 1996; Butt, Fortmann, and Rohrer 1996; Frank and Zaenen 2002). I do not provide a full account of the $\mathcal{D}$-mapping involved in the periphrastic forms in Portuguese, instead focusing on the morphological features. However, my account of some of the periphrastic forms given by the morphological component (section 4.4.9) could be combined with an adaptation of existing LFG accounts of $\mathcal{D}$-mapping. For example, periphrases with *ser* 'be' could include mapping to the feature VOICE: PASSIVE and those with *ter* 'have' to VOICE: ACTIVE, TENSE: PAST and an appropriate value for ASPECT.

Another important division within Portuguese verbal morphology is inflection class. Portuguese verbs are typically divided into three major inflection classes, known as the first, second and third conjugations (abbreviated C1, C2 and C3 in my analysis). Each of these inflection classes is associated with a theme vowel that occurs in the infinitive form of the verb: *-ar*, *-er* and *-ir*. These inflection classes, and their relationships to the Latin inflection classes that preceded them, are discussed at length in Chapter 4. Further 'minor' inflection classes may also be proposed to account for less wide-reaching inflectional patterns, such as those which affect only a small number of lexical items, or a small number of inflected forms within the paradigm. This type of inflection class is used in Taylor's (2012) analysis of Portuguese (see section 3.3.1 below), and also in my analysis of the Portuguese past participles (see Chapter 4).

To clarify the many different distinctions made in the Portuguese verb, I give full verb paradigms for regular verbs from the first, second and third conjugations in Appendix I, including all simple finite forms, and the non-finite forms. I also give the verb paradigms for the auxiliaries *ser* 'be' and *ter* 'have'. In Appendix II I present further information about the periphrastic verb forms with these auxiliaries.

As in any language, Portuguese morphology is closely bound up with phonology. Many well-studied phonological alternations are visible in Portuguese verbal forms, such as the alternation

between close-mid and open-mid front vowels in stressed position in the verbal root in second and third conjugation verbs (e.g. b[e]bo *bebo* 'I drink' vs. b[ɛ]bes *bebes* 'you drink'). These can be treated as morphophonological stem alternations. My focus here is on the interactions of morphology with syntax, rather than with phonology. For this reason, my analysis does not focus on morphophonological alternations. In this respect, I again follow Taylor (2012, 104), who abstracts away from phonological alternations in her analysis of Portuguese morphology, assuming that a separate phonological component accounts for these phenomena. Where I give phonemic transcriptions, these are broad, abstracting away from phonetic detail to capture only the points that are most relevant to a morphosyntactic approach.

Compared with the wealth of inflectional distinctions made in the verb paradigm, nouns in Portuguese have relatively little inflectional morphology, inflecting only for number. Adjectives and determiners inflect for both number (singular or plural) and gender (masculine or feminine). Pronouns inflect for case to a greater or lesser extent depending on the person/number combination represented, with possible distinctions between stressed nominative (subject) and prepositional (oblique) forms, and between accusative (direct object), dative (indirect object) and reflexive/reciprocal clitic forms.[21] However, a full analysis of the morphological properties of all these categories is not the aim of this thesis. In Chapter 6 I consider a phenomenon in the nominal phrase, namely the grammaticalised pronouns in Portuguese that have developed from full noun phrases. This involves a fuller consideration of nominal morphology in Portuguese, particularly in relation to pronouns. Since nominal morphology is more straightforward than verbal morphology, and because I deal with the inflectional morphology of those forms in less depth than the

[21] A much more detailed presentation of the morphology of pronouns, including case, is given in section 6.1.1.

inflectional morphology of the verbs (focusing instead on their syntactic and morphological featural properties), I reserve a more detailed discussion of nominal morphology to that chapter.

Having seen some of the main properties of Portuguese morphology as it is to be treated within PFM, as well as some basic points of analysis of the syntax, I now move on to consider other aspects of the LFG architecture and their role in this thesis.

3.2 Further aspects of LFG

In this section, I look at aspects of the LFG architecture beyond constituent structure, functional structure, and the interface with morphology. I first consider semantic structure, since semantic information is relevant to the lexical entries provided for nouns and pronouns in Chapter 6. I then briefly consider LFG approaches to language change, which are relevant to my accounts of historical change in the morphosyntactic phenomena considered in Chapter 5 and Chapter 6.

Although other areas of LFG theory, such as those developing the modules of information structure and phonological or prosodic structure, may also have useful additional insights into some of the phenomena discussed in this thesis, I do not implement an analysis at these levels of structure, and therefore do not provide details of them here.

3.2.1 Semantic structure

Various approaches to semantic structure have been proposed within LFG. One widely used approach is the *glue theory* of semantic composition (Dalrymple, Lowe, and Mycock 2019, chap. 8), which allows meanings to be assembled by combining logical premises using linear logic (Girard 1987). Linear logic is resource-sensitive, which is to say that meaning contributions are consumed once they are used, and they must be used exactly once in the derivation of the meaning of the whole sentence.

In the glue approach, meaning constructors are used to state how the meanings of individual parts of the sentence combine to provide the meanings of larger phrases or the whole sentence. The

order by which meanings are combined relies mainly on f-structure, rather than being dictated by c-structure order (Dalrymple, Lowe, and Mycock 2019, 289). Meaning constructors are represented as part of the f-description in lexical entries, providing the 'semantic information' mentioned in section 2.3.

A meaning constructor pairs up a meaning (semantic function, given in predicate logic) on the left hand side, with a glue expression (in linear logic) that associates the meaning with a semantic structure on the right hand side:

(57) [**Jane**] *Jane* : j_σ

(58) [**see**] $\lambda x.\lambda y.see(x,y)$: $(\uparrow_{SUBJ})_\sigma \multimap [(\uparrow_{OBJ})_\sigma \multimap \uparrow_\sigma]$

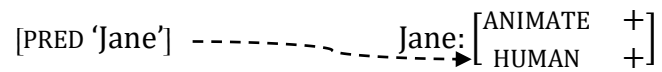
In (58) on the glue side we see the linear implication symbol ⊸. An informal reading of the glue side is 'if we find a semantic resource for the subject and a resource for the object, we can produce a resource for the entire sentence.' (Dalrymple, Lowe, and Mycock 2019, 294). The meaning constructor as a whole can be referred to using its label: examples (57) and (58) above can be referred to as [**Jane**] and [**see**] respectively.

As described by Dalrymple, Lowe and Mycock (2019, 283), semantic functions (the left hand side of the glue expression) can be represented simply by writing their name. For example, the lambda expression in (59) can be represented by the name ***see***.

(59) $\lambda x.\lambda y.see(x,y)$

Using only the name (e.g. ***see***) gives a convenient shorthand for the left hand side of the meaning constructor. The meanings included in the left hand side (semantic function) can also be expanded to provide a greater level of semantic detail. For example, the meaning of the name *Jane* could be expanded and represented as a semantic structure (Dalrymple, Lowe, and Mycock 2019, 313), to show the precise referent and the particular semantic features associated with it. Example (60) shows the meaning constructor for *Jane* with the semantic features +ANIMATE and +HUMAN.

(60) F-structure and meaning constructor for *Jane:*



In this work, I usually provide only a simplified representation of semantic structure, using brief semantic labels such as [**Jane**], to refer to the full semantic structure in lexemic or lexical entries, where full semantic details are not required for my purposes. However, I use semantic structures of the attribute-value matrix type in Chapter 6 to represent the semantic features associated with nouns and pronouns, where more detail is required.

3.2.2 LFG and diachrony

One of the contributions of this thesis is to apply an LFG-PFM perspective to the investigation of diachronic change in morphosyntactic phenomena in Portuguese. In particular, I show that LFG-PFM can provide insights into the fine-grained differences between the grammar that are visible at different stages of a language's history, such as in the development of a mismatch phenomenon, or in changes in inflectional class that affect the realisation of the inflectional morphology of verbs.

The basic perspective of LFG is synchronic: the theory and architecture are developed to allow comprehensive and rigorous analyses of a language at a particular point in time. However, the multi-level approach of the parallel architecture gives clear insights into the independence of different parts of the grammar during diachronic change.

This perspective on LFG is developed notably by Vincent (2001) as part of a volume of LFG work on diachronic phenomena (Butt and King 2001). As Vincent notes, the parallel architecture (with its potential to treat all aspects of language autonomously) allows a more flexible approach to studying historical language data relating to any particular level of structure, particularly including other factors which may be relevant, than an approach in which all facets of the language are modelled using a single set of theory-specific primitives (Vincent 2001, 6).

Existing work shows the advantage of LFG in looking separately at different aspects of the grammar, such as morphology, functional structure and constituent structure. For example, there are LFG approaches to the loss of case marking and the rise of prepositions in marking particular grammatical functions (Vincent and Börjars 2010), as well as in investigations of grammaticalisation, such as from pronouns to case marking (Bresnan and Mchombo 1987; Coppock and Wechsler 2010). To take this last example in more detail, the transition from independent lexical item (pronoun) to functional item (case marking) can be modelled in LFG as involving a loss of the PRED value, and the loss of PERSON and NUMBER feature specifications.

The loss of a PRED feature more generally is seen as characteristic of grammaticalisation processes, since it represents a loss of independent semantic reference. For example, in Chichewa, Bresnan and Mchombo (1987, 775–77) argue that object markers have arisen from pronominal forms which used to have a semantic PRED value 'PRO'. When the forms lose this attribute, they cease to be pronouns, and start functioning as object markers incorporated into the verb, which match the object NP in the grammatical features of person, number and gender. Coppock and Wechsler (2010, 188–89) argue that a possible continuation of this process is the loss of further features, such as person and/or number, as in some Uralic languages. The loss of the person feature and loss of the number feature are independent processes. For example, in Northern Ostyak, verbs agree with the object in number but not in person (Coppock and Wechsler 2010, 191), whereas in Hungarian, they agree with the object in person, but not in number, and there is an additional restriction on topicality (Coppock and Wechsler 2010, 193–96). Examples (61) and (62) show that in Northern Ostyak the same plural suffix *-l* is used on verbs for both third person and second person objects (Coppock and Wechsler 2010, 188), even though generally (and historically) verbal agreement usually involves person.

(61) Ma täm kälang wel-sə-l-am
I this reindeer kill-PAST-PLOBJ-1SGSUBJ
'I killed these reindeer.'

(62) Xŭnśi näng mŭng-iluw xälśa want-lə-l-an?
when you we-ACC where see-PRES-PLOBJ-2SGSUBJ
'When did you see us where?'

Coppock and Wechsler provide the summary given below as Table 6 (2010, 189) for a situation where the person feature is lost, as in some Finnish varieties (Toivonen 2001) as well as Northern Ostyak.[22]

Stage 1: Pronoun	**Stage 2: Loss of reference**	**Stage 3: Loss of person**
(↓PRED) = 'pro'	~~(↓PRED) = 'pro'~~	~~(↓PRED) = 'pro'~~
(↓INDEX PERS) = *p*	(↓INDEX PERS) = *p*	~~(↓INDEX PERS) = *p*~~
(↓INDEX NUM) = *n*	(↓INDEX NUM) = *n*	(↓INDEX NUM) = *n*
(↓INDEX GEND) = *g*	(↓INDEX GEND) = *g*	(↓INDEX GEND) = *g*

Table 6 Stages of lexical specification in the transition from pronoun to agreement

Proposals in this area are discussed further in Chapter 6, when I consider the grammaticalisation of pronouns from nouns.

Within the LFG architecture it is possible to develop multiple separate, incrementally distinct representations of fine details of the grammar as it undergoes change, as seen in the example above,

[22] INDEX is a bundle of agreement features including PERSON, NUMBER and GENDER, which can be specified separately from and contrasted with CONCORD features. These concepts are discussed further and implemented in Chapter 6: see section 6.1.3.

where the loss of the PERSON feature is independent from the change in function from pronoun to agreement captured by the loss of the PRED attribute. This also presents the possibility of representing divergence, as different language varieties develop in different ways. I develop these ideas throughout the rest of this thesis, showing that a fine-grained approach that takes each aspect of the grammar on its own terms gives clear insights into the possibilities of diachronic divergence within Portuguese.

In the next section, I move back from considering LFG to looking at PFM. More particularly, I discuss extensions to the original theory of PFM, and consider their relevance to this work.

3.3 Extending PFM

In this section, I describe some extensions to the basic theory of PFM, and explain how and why some of these are employed in later chapters of the thesis to provide full analyses of the Portuguese data under consideration.

3.3.1 Taylor's (2012) adaptations of PFM

As described above, Taylor (2012) presents a detailed comparative analysis of Romance verbal morphology in PFM, including the inflectional paradigms of Portuguese verbs. Her work presents some developments of the basic PFM theory, particularly in regard to stems. As described above, PFM and LFG allow effective treatments of stems as crucial parts of inflectional morphology: in PFM (section 2.2), the stem selection rules ensure the correct stem is selected, while in LFG, as we saw in section 2.3.5, idiosyncratic morphological information, including stems, is stored in the lexemic entry. This enables the two theories to account for the morphology of languages with stem alternations (section 2.2).

Taylor emphasises the 'morphomic' nature of stems: that is to say, the fact that stems may be selected arbitrarily, rather than associated with morphosyntactic property sets. Above we saw the example of the Spanish PYTA stem, which occurs in several otherwise unrelated verb tenses.

Stems are in this sense not meaningful. In addition, Taylor presents clear evidence from the Romance languages of complex stems: these are stems formed on the basis of other stems, which act independently as the basis for fully inflected forms. An example of a complex stem is the Spanish *r*-stem (e.g. *cantar* for 'sing'), which is formed from the basic stem + theme vowel (e.g. *canta*), with additional phonetic material ('r'). The complex stem is used in the formation of the future and the conditional, and for nearly all verbs, is also the infinitive (e.g. *cantar* 'sing.INF', *cantará* 'sing.FUT.1/3SG', *cantaría* 'sing.COND.1/3SG').

To account for the existence of complex stems, Taylor proposes the existence of a stem formation function (abbreviated SF) which is similar to the paradigm function (PF) described in section 2.2.1. The stem formation function allows the selection of the appropriate stem root (in its first rule block), and the addition of further phonetic material (in subsequent rule blocks) where required, for a complex stem. For a basic stem, the stem root is selected, and later rule blocks do not change the form. Taylor argues that it is necessary to make a distinction between the rules that form stems, and the standard realisation rules. This is because realisation rules are reserved to provide phonetic strings 'as exponents of morphosyntactic properties' (C. Taylor 2012, 228), but following Spencer (2012), all stems are morphomic, and are not exponents of morphosyntactic properties. In this respect, Taylor builds on two crucial insights into the identification and formal treatment of stems, presented by Spencer (2012):

(63) Strictly Morphomic Stem Hypothesis: Assume all stems are morphomic (thereby simplifying form ⇒ function mappings)

(64) Maximizing Stems Principle: Assume all recurrent partials are (part of) stems unless they can be shown to be exponents of inflectional (morphosyntactic) properties

These two principles allow the identification of both simple and complex stems, as described briefly by Spencer (2012) and in more depth by Taylor (2012).

In Luís' earlier approach to Portuguese morphology, the concept of a complex stem is not explicitly mentioned, although similar insights about the nature of the morphomic stem and possibilities for its formation are incorporated (Luís 2004, 176, 351–52). Luís considers the future, conditional and imperfect tenses of Portuguese verbs, and suggests that at least two stems are required to form these tenses: $stem_1$ which consists of root + theme vowel[23] (e.g. *leva* for *levar* 'carry') and $stem_2$ which consists of $stem_1$ + -r (e.g. *leva+r*) (Luís 2004, 170–72). However, the precise relationship between the two stems, particularly regarding how $stem_2$ is formed in relation to $stem_1$, is left for future research. This gap is effectively filled by Taylor's detailed presentation of stem formation rules. Essentially, in these rules, a basic stem is formed through one iteration of stem selection and formation rules, and a second iteration of the same rule blocks then adds further phonetic material and a second theme vowel as required.

For example, to form the Portuguese future tense verb *levará* 'he/she/it will sing', the theme vowel *a* is added to the basic stem root *lev*, then in the second iteration of the rules, the consonant *r* and a null theme vowel are added,[24] to give the complex stem *levar*. The realisation rules then add the additional inflectional ending *á* which realises the third person singular for the future indicative, giving the final form *levará* as desired. Further details on the selection and realisation of complex stems in Taylor's analysis and my development of it are given in section 4.4.5.

[23] The word 'theme' or 'thematic' in this context refers to a word's stem.

[24] A null theme vowel is required because in Taylor's analysis, stems are distinguished from roots by the presence of a theme vowel (whether phonetically overt or null).

An important part of Taylor's model is that each verb has a number of stem roots (one or more), stored as an array[25] in the lexemic entry.[26] The verb's theme vowels are also stored in the lexemic entry as an array. From these stem roots and theme vowels, stems can be formed: a stem is defined as a stem root (sometimes with additional phonetic material) suffixed with a theme vowel. In the example of *levará*, the basic stem *leva* has the basic stem root *lev* suffixed with the theme vowel *a*, and the complex stem *levar* has the basic stem *leva*, additional element /r/, and a phonetically null theme vowel.

The concept of the theme vowel in Taylor's model is therefore different from in other approaches to morphology. The relevant defining feature of the theme vowel here is that it is combined with a stem root to give a stem, whether basic or complex. There are some differences between the two: the theme vowels used to form basic stems vary from verb to verb, such as by inflection class properties (for example, theme vowel 1 for first conjugation verbs is /a/), while the theme vowels used to form complex stems are generally constant across all verbs. By contrast, in standard approaches to Romance morphology, only the vowel used to form the basic stem would be recognised as a theme vowel, and there is therefore a strong association between the theme vowel and the conjugation class (/a/ for first conjugation, /e/ for second, and /i/ for third). For

---

[25] Taylor defines arrays as ordered 'containers' which may contain the same value more than once. In these respects, they are different from sets, which are unordered containers which can contain a particular value only once (C. Taylor 2012, 387). Arrays are thus similar to 'lists' in some computational settings, although lists can contain values of different types.

[26] Taylor refers to the 'lexeme's lexical entry': in the terminology used in this thesis, this corresponds to the lexemic entry, as it clearly means the information associated with the whole lexeme (2012, 229).

example, in their brief summary of Portuguese verb morphology, Dubert and Galves (2016, 424) note that 'in the second and third persons singular of the present indicative … the thematic vowels of the second/third conjugations are neutralised as [ɨ] in European Portuguese and [i] in Brazilian Portuguese'. Vowels that are identified as thematic in Taylor's analysis, such as the final vowel [o] in the Portuguese gerund forms (e.g. *cantando* 'singing': this non-finite form is treated as a complex stem), would not be seen as thematic in standard Romance morphology: only the last vowel of the basic stem *canta* would count as a theme vowel, because the full form is not recognised as a stem, and therefore does not end in a theme vowel.[27] In this work, I maintain Taylor's characterisation of a theme vowel as being the final vowel in a stem.

Another crucial aspect of Taylor's approach to verbal morphology is *re-indexing*. This is a process of relating more complex patterns of stem root selection to simpler ones, allowing a pattern to be discerned for each language as to where the different stem roots are used, and ultimately identifying the basic stem for each lexeme (C. Taylor 2012, 147–49). Verbs with a more irregular schema can

[27] Theme vowels play a different role in Distributed Morphology (DM): they are treated as 'exponents inserted into Theme positions (TH) … [TH nodes] are added to *v* (and other functional heads) at PF' (Embick and Halle 2005, 11–12). (PF here stands for phonetic or phonological form. This determines phonological form without affecting semantics.) The TH node is 'spelled out' with a theme vowel according to the conjugation class feature of the root, thus the theme vowel acts as an indication of the conjugation class. The TH position can also be deleted to account for forms which do not have a theme vowel. In DM, there is no concept of a stem, so there is no relationship between the selection of a theme vowel and the formation of a stem. I refer the reader to Taylor (2012, 74–76) for a refutation of Embick and Halle's arguments against the existence of stems.

be reindexed to verbs with a more regular schema. Taylor gives the example of the present indicative of the following Italian verbs, which are shown with the distribution of stems across person/number combinations from most irregular to most regular (C. Taylor 2012, 148).

a. DOLERE 'hurt': T1 = {2pl}, T2 = {1sg,3pl}, T3 = {2sg,3sg}, T4 = {1pl}
b. VENIRE 'come': T1 = {1pl,2pl}, T2 = {1sg,3pl}, T3 = {2sg,3sg}
c. VALERE 'be worth': T1 = {2sg,3sg,1pl,2pl}, T2 = {1sg,3pl}
d. CANTARE 'sing': T1 = {1sg,2sg,3sg,1pl,2pl,3pl}

This allows the stem in T1 to be identified as the basic or default stem in Italian. This process allows a convenient way of relating stems: for example, we can state in which circumstances stems are used by irregular verbs, and generalise that they are reindexed to default stems in more regular verbs.

A further adaptation of PFM that Taylor proposes, which I adopt without further developments, is altering the Paradigm Function so that it takes as input and produces as output an array of forms (rather than a single form). This accounts for instances where a single paradigm cell can be filled by two alternative forms, with no morphosyntactic or semantic reason for their distribution, such as the Spanish imperfect subjunctive forms e.g. *cantase* and *cantara* 'sing.IPFV.SBJV.1/3SG'. Taylor adapts the standard PFM definitions to represent arrays as follows (C. Taylor 2012, 226):

(65) Format of a stem selection rule (SSR):

$SSR_{n,\tau,C}$ (<χ, σ>) =def <ψ', σ>

where ψ' is an array of potentially morphophonologically modified phonetic strings

(66) Format of a realisation rule (RR):

$RR_{n,\tau,C}$ (<χ, σ>) =def <ψ', σ>

where χ is an array of phonetic strings and ψ' is an array of potentially morphophonologically modified phonetic strings

In this notation, n represents the rule block index, τ represents the morphosyntactic property index, and C represents the class index.

The morphological model used by Taylor is complex, involving much lexeme-specific information as well as many language-general morphological rules and principles, enabling the full range of morphological behaviour to be captured. The system's complexity is reduced by principles of defaults and inheritance. For example, Taylor describes a generic lexical (= lexemic) entry for each word class, which exists so that information shared across lexemes can be included here, rather than specified in the lexemic entry of each individual lexeme (2012, 304–5). Such information could include a default rule for the formation of stems from a stem root and theme vowel (the 'lexical stem rule'), and perhaps other information like default theme vowels. This concept has similarities with systems of defaults or inheritance used in other linguistic theories, such as the type hierarchies used in Head-Driven Phrase Structure Grammar (HPSG) (Pollard and Sag 1994), and the templates used in LFG to represent functional information, whether within the lexical entry, or used as annotations on phrase structure rules (see for example Dalrymple, Kaplan, and King 2004; Asudeh, Dalrymple, and Toivonen 2008; Dalrymple, Lowe, and Mycock 2019, sec. 6.8).[28]

[28] There are important differences between HPSG type hierarchies and the template hierarchies used in LFG: HPSG hierarchies involve strict inheritance, while the LFG template hierarchies represent different kinds of relationships between f-descriptions, such as conjunction or negation (Dalrymple, Lowe, and Mycock 2019, 233).

For the sake of making my LFG analyses explicit, I do not make use of LFG templates, and do not discuss further the possible relationship between this concept, and Taylor's concept of a generic lexical entry shared across verbs. However, it is clear that Taylor's PFM approach has much in common with LFG templates in terms of capturing information shared across items in the lexicon, albeit at the morphological rather than the syntactic level.

Taylor uses the following features for verbal inflection:

- Mood (md): indicative, subjunctive, imperative
- Tense (including aspect) (tns): past 1 (= 'imperfect'), past 2 (= 'perfect', 'simple past' or 'preterite'), present, future, conditional
- Agreement (agr) including Person (per): 1, 2 ,3 and Number (num): sg, pl
- Vform (for non-finite forms): past participle (ppl), gerund (ger), infinitive (inf)

Tense and aspect are combined in this analysis because there is little evidence of systematic aspectual contrasts across verbal inflection in Romance languages in general, or Portuguese in particular. The two past tenses usually called 'imperfect' and 'perfect' or 'preterite' are therefore abbreviated as past 1 and past 2, and are not treated as distinguished by aspect.

Voice is not treated as a feature in Taylor's analysis of Romance languages, since she deals only with simplex (one-word) forms rather than with periphrases, and does not consider the active versus passive properties of past participle. However, she does treat voice as a feature when considering Latin, which has a simplex passive. In my analysis, I use the feature of voice with two values (active and passive): see section 3.3.4 for discussion.

More features of Taylor's PFM proposals are discussed in Chapter 4, where I build on her analysis of Portuguese to account for the double past participle phenomenon. I now move on to consider another extension of PFM: Spencer's (2013) Generalized Paradigm Function Morphology

(GPFM), which plays a role in Chapter 4 in accounting for the properties of past participles in relation to verbal paradigms.

3.3.2 Generalized Paradigm Function Morphology (Spencer 2013)

Spencer (2013) extends the principles of PFM to account for the many different ways in which lexical items are related, distinguishing not only canonical inflection and canonical derivation, but also other intermediate types of lexical relatedness. His Generalised Paradigm Function Morphology (GPFM) theory gives an insight into morphological relationships that are otherwise hard to account for in solely inflectional models of morphology. The core technical proposal that Spencer makes is the 'Generalised Paradigm Function' or GPF (Spencer 2013, 178):

(67) $\text{GPF}(\langle \text{LEXEME}, \sigma \rangle) =_{\text{def}} \langle \text{LEXEME}' \rangle$

In this function, both LEXEME and LEXEME' are four-dimensional lexical representations with the attributes FORM, SYN, SEM and LI (lexical index),[29] and σ is a set of features. The GPF has four components, each of which applies to one of the attributes: f_{form}, f_{syn}, f_{sem} and f_{li}. The GPF accounts for the relatedness of lexemes, whether inflectional, derivational, or in any of the twelve intermediate categories defined by the identity or distinctness of the attributes FORM, SYN, SEM and LI. In Spencer's model, when comparing two given lexemes, each of these features can be either distinct or identical. This gives sixteen possible combinations, of which one is logically impossible (LI is different; FORM, SYN and SEM are the same) and one is 'identity' (all four are identical). Of the remaining fourteen, one is canonical derivation and one is canonical inflection, leaving twelve intermediate, hypothetically possible types, which Spencer claims are all instantiated, and some of which are extremely common. These different categories show a great range of behaviours,

[29] These are similar to the four elements of the lexemic entry described by Dalrymple (2015): see Chapter 2.

necessitating a more nuanced approach to lexical relatedness and category properties than that permitted by a simple 'inflection or derivation' view of morphology.

The different possibilities are set out in Table 7. In the table, the minus sign indicates that the two features are identical for the two lexemes, and the plus sign indicates that they are distinct. The table brings together information presented by Spencer (2013, 7–8, 273, 275: Tables 6.8 and 6.9).

	FORM	**SYN**	**SEM**	**LI**
Default representation (identity)	-	-	-	-
Contextual inflection	+	-	-	-
M-inert transposition	-	+	-	-
Transposition	+	+	-	-
Weak polysemy	-	-	+	-
Inherent inflection, evaluative morphology	+	-	+	-
Angelstellte(r) nouns	-	+	+	-
Meaningful transposition	+	+	+	-
Logically impossible	-	-	-	+
Perfect synonymy	+	-	-	+
M-inert transpositional lexeme	-	+	-	+
Transpositional lexeme	+	+	-	+
Strong polysemy	-	-	+	+
Syntactically inert derivation	+	-	+	+
M-inert derivation	-	+	+	+
Canonical derivation	+	+	+	+

Table 7 Types of lexical relatedness

GPFM makes use of an extensive default-based system, similar to mechanisms in Head-Driven Phrase Structure Grammar (HPSG: Pollard and Sag 1994), by which properties that are not defined by the GPF are inherited from the base lexeme. Depending on the type of relatedness at issue, different properties will be inherited. For example, we can consider the second row in Table 7, pure (contextual) inflection: in the relationship of the English third person singular present form

writes to the base lexeme *write*, only FORM is altered, and SYN, SEM and LI are inherited from the base lexeme.

Spencer gives a principled answer to the question of whether two lexical items are in the same paradigm or not: items that share a LI (such as inflected forms) are members of the same inflectional paradigm, while those that have a different LI are not in the paradigm. Derived forms have a different LI from their base lexeme, and are part of a two-cell 'derivational paradigm', which consists of the base lexeme and the derived form. For example, the English verb *write* has a derived subject-nominal form *writer*. This general pattern of derivation is found in many English verbs (Spencer 2013, 164–68). The derivational paradigm could be schematised as in (68):

(68)

write	writer

The Generalised Paradigm Function applies to the form *write* to give the derived subject nominal form *writer* as follows (Spencer 2013, 178–79):

(69) GPF(<WRITE, SN>) ≡

f_{form} (<WRITE, SN>) = *write er*

f_{syn} (<WRITE, SN>) = Noun

f_{sem} (<WRITE, SN>) = [PERSON(x), [WRITE(x,y)]]]

f_{li} (<WRITE, SN>) = WRITER

In canonical derivation, the last row of Table 7, the four attributes of the base lexeme – FORM, SYN, SEM and LI – are all changed.

This theory is particularly useful for investigating instances where the relationship between lexical items is unclear: for example, are we looking at two forms of the same verb, or is one of the forms a separate lexical item? In this thesis, I generally consider canonical inflection. However, I use some GPFM concepts in my discussion of past participles in Chapter 4, since the relationship between verbs, past participles and related adjectives is more complex than pure inflection.

3.3.3 PFM2 (Stump 2016)

The theory sometimes known as PFM2, presented in Stump (2016), adapts and extends many of the mechanisms introduced in standard PFM (that is, Stump 2001). Although I do not implement PFM2 proposals in this thesis, here I give some details of the differences between PFM2 and the original PFM theory, which provide useful context for my combination of PFM with LFG, as well as for understanding other developments of PFM.

The most crucial difference is that PFM2 introduces three distinct paradigms: the content paradigm <LI, σ>, the form paradigm <S, τ>, and the realisation paradigm <R,τ> (Table 8). Example cells are taken from Stump (2016, 105, Table 7.1).

Paradigm	**Is a pairing of**	**Example cell**
Content paradigm	lexemes and properties	<SING, {3sg prs ind}>
Form paradigm	stems and properties	<sing, {3sg prs ind}>
Realisation paradigm	inflected forms and properties	<sings, {3sg prs ind}>

Table 8 Paradigms in PFM2 and their cells

Cells in content paradigms are pairings of lexemes with properties (morphosyntactic features); cells in form paradigms are pairings of stems and properties (including morphomic features as well as morphosyntactic features); and cells in realisation paradigms are pairings of inflected forms and properties.

The correspondence function *Corr* from the content paradigm to the form paradigm is made up of two functions: *Stem* and *pm*. *pm* is the 'property mapping function', which applies to the set of morphosyntactic features to be realised (σ) to give this set of properties plus morphomic features (τ), while *Stem* 'applies to the cell's lexical and morphosyntactic content to yield the stem form Y

employed in its realization: *Stem*(<L,σ>) =Y' (Stump 2016, 77).[30] *PF* is the Paradigm Function, familiar from the simpler PFM1 architecture, but which here has a different function, mapping the form paradigm to the realisation paradigm. Biswas (2019) introduces terminology to distinguish *PF* from *PFc*-r which is the mapping from the content paradigm to the realisation paradigm (that is, the combination of *Corr* and *PF*).

Figure 4 illustrates this architecture (based on Biswas 2019, fig. 4). The uppermost box in the figure represents the link with the LFG architecture (Figure 2), using terminology from LFG: LI is the

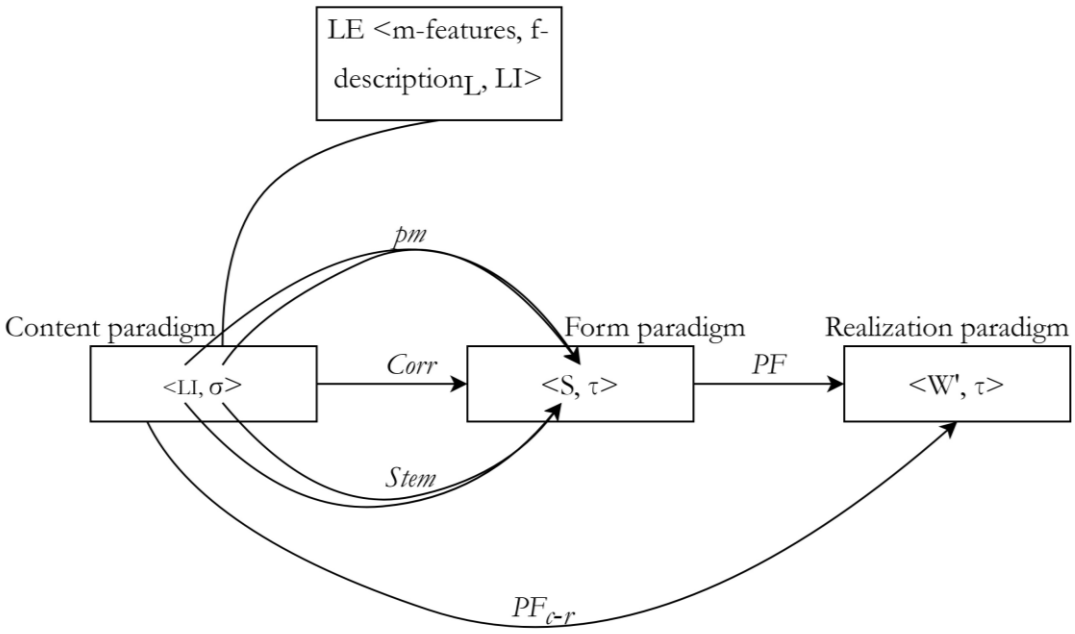


Figure 4 Schematic representation of PFM2 architecture

[30] The *Stem* function in PFM2 and the stem selection rules in Taylor's adaptation of PFM have some similarities, since they both operate as part of the overall Paradigm Function to provide the correct stem as the basis for a fully inflected form, given a set of properties. However, in Taylor's model, there is no separate 'form paradigm' consisting of pairings of stems and properties: the correctly formed stem feeds directly into the subsequent realisation rules to give the inflected form.

Lexemic Index and LE the lexemic entry. This reflects that information relevant to the whole lexeme (e.g. m-features) contributes to the content paradigm, giving input to produce the correct realised (inflected) form.

Some abbreviations used in the figure are specific to Biswas's presentation of PFM2: for example, Stump (2016) does not use the name $PF_{c\text{-}r}$ (instead this is *PF* in his presentation), or the abbreviation W (this is Stump's *R*, which Biswas alters to avoid confusion with the LFG abbreviation $\mathscr{R}$). The figure summarises much of the information contained in the diagrams and tables given by Stump (2016, 114–15).

The role of the LFG description function $\mathcal{D}$ overlaps in some ways with the PFM rules of paradigm linkage. To be more precise, $\mathcal{D}$ corresponds to the property mapping function (which relates the content paradigm to form paradigm), which is one of several 'rules of paradigm linkage' where disruptions in the canonical links between paradigms can occur.

In PFM2, the default rule of paradigm linkage (Stewart and Stump 2007, 392) is:

(70) Where R is L's root, <L, σ> → <R, σ>

In this default rule, a single morphological root R is associated with the set of properties σ. There are many instances of this default rule being interrupted, for example in deponency[31] or syncretism.[32] In some lexemes (such as Latin deponent verbs, which have passive form but active

[31] For further information on deponency, I refer the reader to Baerman et al. (2007). For an account of the morphological typology of deponency, including the analysis of deponency as a mismatch between form and function, see Baerman (2007a).

[32] For an overview of syncretism, see Baerman, Brown and Corbett (2005) and Baerman (2007b).

meaning), different stems are associated with different sets of properties, and there are regular correspondences such as:

(71) $\langle L, \sigma \rangle \rightarrow \langle X, \sigma \rangle$

$\langle L, \sigma' \rangle \rightarrow \langle Y, \sigma' \rangle$

In syncretism, there is a correspondence between the forms of two cells in the paradigm. Syncretism is defined by Matthews (2007) as:

> 'The relation between two or more words in a paradigm that have different morphosyntactic features but are identical in form. Thus the distinction between a past tense and a past or passive participle is syncretized in the paradigms of most verbs in English: past tense *walked* and participle also *walked*; past tense *won* and participle also *won*; and so on. … Used especially when the identity is regular across all paradigms.'

In PFM2, syncretism is captured by cells with different property sets (σ) in the content paradigm mapping to the same cell ($\langle S, \tau \rangle$) in the form paradigm. The possibility of capturing syncretism using rules of paradigm linkage in PFM2 means that rules of referral, a key component of PFM1, do not play a role in PFM2. As noted by Stump (2016, 258), in PFM1, rules of referral and rules of exponence are both types of realisation rules that together project inflected wordforms from the stem, but in PFM2, 'instances of systematic syncretism are … defined as part of the mapping from content cells to their form correspondents.'

In the LFG interface between syntax and morphology discussed above, $\mathcal{D}$ establishes the relationship between morphological features and their syntactic correlates, so it is also the point at which mismatches between these relationships, where morphological features do not align straightforwardly with their corresponding syntactic properties, are captured. Deponency is a good example of this, as set out by Dalrymple, Lowe and Mycock (2019, 462–63). The rules given for

Latin deponent verbs establish a divergence from the usual correspondence between passive morphology and passive syntax (Dalrymple, Lowe, and Mycock 2019, 463):

(72) $\mathcal{D}$-mapping dependent on the M-CLASS m-feature:

D_{feat}<LI, M-VOICE:PASS, m-features, *f*, {(*f* VOICE) = ACTIVE}>

if and only if M-CLASS:DEPONENT-VOICE ∈ m-features.

D_{feat}<LI, M-VOICE:_VOICE, m-features, *f*, {(*f* VOICE) = _VOICE}>

if and only if M-CLASS:REGULAR ∈ m-features.

In these $\mathcal{D}$-mapping rules, if a verb is in the morphological class 'DEPONENT-VOICE', then the morphological voice feature 'PASSIVE' will correspond to the syntactic voice feature 'ACTIVE'. By contrast, for other verbs (those in the REGULAR class), there is no unusual correspondence between morphological and syntactic voice features: the default correspondence between M-VOICE and *f* VOICE will be used.

This same phenomenon, of Latin deponent verbs, is treated in detail both in the PFM2 theory (Stump 2016, 197–202), and in the syntax-morphology interface described by Dalrymple, Lowe and Mycock (2019, 462–63). This points to an overlap or redundancy in the two models: both approaches are capable of resolving complex problems in morphosyntax, but from different perspectives, with the PFM2 approach dealing with more phenomena within an extended concept of the morphology, and with LFG putting more emphasis on the mapping between morphology and syntax. The PFM2 approach goes well beyond typically morphological problems, to account for more syntactic phenomena in ways that are not required when working with the LFG architecture, which already includes mechanisms to deal with these. Compared with PFM1, PFM2 introduces considerably more complex machinery, which has clear advantages in accounting for languages with particularly intricate patterns of stem distribution, such as Sanskrit, as demonstrated in detail by Stump (2016).

However, as can be seen from the shared example of Latin deponent verbs, there are redundancies in attempting to combine the LFG architecture with the three paradigms of PFM2. I therefore use the original PFM (Stump 2001) to provide the inflectional morphological component required. This does not rule out future work to develop a coherent integration of PFM2 theory with the LFG architecture.

3.3.4 Periphrasis in PFM (Biswas 2019)

Another direction in which the theory of PFM has been extended is in accounts of inflectional periphrasis: that is, when a single cell in the inflectional paradigm is filled by a combination of two words. A more technical definition of periphrasis is given by Spencer (2013, 103): 'a construction involving two or more syntactic terminals (syntactic words) which expresses a purely grammatical property, that is, a property which would be inflectional if it were expressed morphologically.' In Portuguese, as we saw above, verbal periphrases typically consist of an inflected auxiliary and a non-finite form (infinitive, gerund or past participle). Portuguese verb forms with *ter* 'have', which are associated with particular aspectual properties (see section 3.1.2), can straightforwardly be regarded as periphrases, which can be derived within the inflectional paradigm. Given the parallelism between these forms and passives formed with *ser* 'be', I propose to extend the same analysis to the passives, even though there is no simplex inflectional morphological realisation of voice features in Portuguese. This follows Spencer's (2013, 103) analogous argument about the English perfect and passive, i.e. 'I have eaten' vs. 'It was eaten'. English has no morphological passive, but the parallelism between the two structures justifies the treatment of both as

periphrases within the paradigm. In my analysis, the active and passive periphrases are distinguished by the presence of a syntactic VOICE feature with the value ACTIVE or PASSIVE.[33]

Spencer (2013, 221–22) outlines requirements for a PFM account of periphrases of this sort, with realisation rules that specify a periphrastic realisation, giving the example of the English progressive and perfect aspects:

> 'the rules would state that [ASPECT:perfect], say, is realized by a syntactic template which includes the *have*-auxiliary followed by the perfect participle of the next verb in the syntactic sequence … Part of the definition of the periphrasis would then be the specification of the content paradigm for the two components of the periphrasis.'

Various approaches to periphrasis are possible within LFG. Most relevant to this work, Biswas (2017) presents an account of periphrasis in PFM and LFG, developed further in Biswas (2019). This account introduces collocational requirements into the morphology of paradigm cells realised by periphrases: the appropriate auxiliary verb form appears alongside the past participle or other non-finite form. The theoretical details of these collocational requirements are presented here, along with an indication of how these can be adapted to work alongside the rest of my theoretical architecture.

Biswas's (2019) account of periphrasis builds on PFM2. As described above, in PFM2, the normal mapping between a cell in the content paradigm and a cell in the realisation paradigm takes place through the function $PF_{c\text{-}r}$. This is captured in (73):

(73) $PF_{c\text{-}r}<LI, \sigma> = <W, \tau>$

[33] See section 4.4.10 for further information on the syntactic and semantic treatment of voice in relation to passive periphrases.

In periphrasis, this normal mapping is overridden, by adding an element ω which applies collocational constraints:

(74) $PF_{c\text{-}r}$<LI, σ> = < <W, τ_1> ω>

Where W is a wordform and ω is the collocational constraints that W must satisfy: W and ω taken together constitute the realisation of σ.

This essentially means that the realisation of the pairing <LI, σ> is the correct wordform and the collocational constraints needed.

In Biswas's account, ω is 'a list of collocational constraints that will be mapped to m-structure by actions of $\mathcal{D}$ [the description function] and $\mathscr{L}$ [the lexical entry]', although he suggests this implementation could be simplified (2019, 17). An example of the collocational constraint proposed for the French *passé composé*, a past tense periphrasis formed with either the auxiliary *avoir* 'have' or *être* 'be' and the past participle, is as follows (Biswas 2019, 17):[34]

(75) If TENSE:passécomposé ∈ σ & AUX-SEL: avoir-aux ∈ m(LI),

ω = <{M-AUX-C:+, M-LID-C:AVOIR-AUX, M-TENSE-C:PRES, M-DEP-AUX:-, M-DEP-VFORM:past-part},

σ U {SUBJPERS:_, SUBJNUM:_}>

This can be paraphrased as follows: if σ (the set of features to be realised) contains the tense *passé composé*, and the morphological features of the lexeme include the feature 'select auxiliary *avoir*', then the collocational requirement (ω) is that there must be an auxiliary, there must be a form of the auxiliary *avoir*, its tense must be present, and there must be a dependent which is not an

[34] M-LID refers to the arbitrary lexical identifier of the auxiliary, i.e. AVOIR-AUX is the LID for the auxiliary verb *avoir*, as distinct from the lexical verb *avoir*.

auxiliary, and it must be in the past participle form; and the σ (feature set) of this auxiliary also must contain the person value and number value of the subject. Features with an m-prefix such as M-LID-C:AVOIR-AUX and M-DEP-AUX are mapped to an m-description, and features with a c-suffix such as M-LID-C:AVOIR-AUX are mapped to a constraining equation.

This collocational requirement is introduced in PFM2 by the more general principle given below for the *passé composé* (Biswas 2019, 15):[35]

(76) If TENSE:passécomposé ∈ σ & AUX-SEL: avoir-aux ∈ m(LI),

$PF_{c\text{-}r}$<LI, σ> $=_{def}$ < $PF_{c\text{-}r}$<LI, (σ/VFORM :past-part −{TENSE:_, MOOD:_}) >, ω>

This gives the lexical entry below (Biswas 2019, 16) for the past participle *couru* 'ran'. The effect of the collocational requirements (ω) is seen in the introduction of m-descriptions, where the f-structure designator is annotated with $_{\mu}$.

[35] The bracketed part means 'replace the relevant feature in σ with the feature "verb form: past participle" and remove any tense or mood features'.

(77) *couru,* V[main,part],

(↑PRED) = 'run<SUBJ>'

(↑VFORM) = FIN

(↑TENSE) = PASSÉCOMPOSÉ

(↑SUBJ PERS) = 3

(↑SUBJ NUM) = SG

$(\uparrow_{\mu}$DEP AUX) = −

$(\uparrow_{\mu}$DEP VFORM) = PAST-PART

$(\uparrow_{\mu}$VFORM) $=_c$ FIN

$(\uparrow_{\mu}$AUX) $=_c$ +

$(\uparrow_{\mu}$LID) $=_c$ AVOIR-AUX

$(\uparrow_{\mu}$TENSE) $=_c$ PRESENT

$(\uparrow_{\mu}$SUBJ PERS) $=_c$ 3

$(\uparrow_{\mu}$SUBJ NUM) $=_c$SG

I propose that collocation requirements could more straightforwardly be introduced in the original PFM architecture, where only one function (the Paradigm Function) applies to the representation of a lexeme with the relevant properties (<L, σ>) and the realisation in a pair with these properties (<R, σ>). The collocation requirements (ω) could then be added on to the realisation cell as <<R, σ> ω>. This results in a simpler analysis overall than one which uses PFM2, while retaining the necessary collocation requirements associated with the realisation of periphrases.

In Chapter 4, I describe how this proposal can account for Portuguese periphrastic verbs, giving worked examples for the compound pluperfect and present passive forms. It is worth noting that

the requirements for Portuguese are simpler in two respects than those for French.[36] First, French verbs are lexically specified to take either *avoir* or *être* as their auxiliary in the *passé composé*, but in Portuguese, the selection of a particular auxiliary verb depends on the voice of the periphrasis.[37] Active periphrases use the auxiliary *ter* and passive compound forms use the auxiliary *ser*. The lexical entry for Portuguese verbs therefore does not need to include any mention of auxiliary selection. Second, past participle agreement is simpler in Portuguese than in French: past participles agree when used in passive forms (with *ser*) and do not agree in active periphrases (with *ter*).

Since I explore the complex behaviour of some Portuguese periphrastic forms, namely those involving past participles, in Chapter 4 and Chapter 5, it is important to have a working account of periphrasis. However, the account described here is just one possibility within the LFG-PFM architecture. One analysis may be more suitable for one language, and a different one for another: trivially, some distinctions in analysis will be required to treat French periphrases as compared with Portuguese ones, due to the differences described above. Alternative analyses building on the PFM morphological model may also be implemented, with little effect on the wider architecture, demonstrating the flexibility of the LFG architecture for accounting for morphosyntactic phenomena. Both PFM and PFM2 analyses, as well as analyses not using PFM, would be able to account for the data satisfactorily, with some trade-offs in perceived efficiency, simplicity and

[36] These differences between the two languages do not mean that only a PFM2 analysis is appropriate for French. A PFM(1) analysis of the French data would also be possible.

[37] For the constructions considered here: other active periphrases involve auxiliaries such as *estar* 'to be' and *ir* 'to go', contributing different tense and mood features (see section 3.1.2).

redundancy. My aim here is to demonstrate that PFM represents an effective and non-redundant option for analysing Portuguese verbal morphology, including periphrastic forms.

3.4 Chapter summary

This chapter has continued to develop the LFG-PFM framework used throughout the rest of the thesis, in particular looking at more complex areas of the theory and considering their application to Portuguese data. We have seen that LFG and PFM can be combined and extended to cover a wide range of morphological and syntactic phenomena, a property that is useful for analysing many kinds of complex grammatical behaviour.

In the subsequent chapters of this thesis, I apply the theories in more depth to analyse two phenomena from the morphology and syntax of Portuguese which involve mismatches. The first of these, in the next two chapters, is the double past participle phenomenon. As we shall see, having recourse to a powerful theory of inflection, with close links to the syntax, is vital in giving a full account of this phenomenon.

Chapter 4 Portuguese double past participles in synchrony

This chapter investigates Portuguese verbs with two distinct past participle forms that are used in different syntactic contexts. Although the phenomenon can be described in simple terms, the existence of double past participle forms raises many interesting questions about the interactions between morphology and syntax, and about how the different past participle forms are best analysed in a formal model of the grammar.

The aim of this chapter is to present a theoretical analysis using LFG and PFM that accounts for the morphological and syntactic phenomenon of double past participles, investigating the issues involved in this analysis. These include:

- the relationship between form and function in past participles
- the lexical category membership and related grammatical behaviour of past participles (verbal and adjectival)
- the position of past participles as non-finite forms in the verbal paradigm
- the role of the past participle in verbal periphrases, including the passive

In the rest of section 4.1, I introduce past participles in Portuguese, considering their different functions (section 4.1.1) and the forms they can take (section 4.1.2). I then consider the lexical category of past participles, namely their status as verbal and/or adjectival forms (section 4.2), looking at several criteria which have been proposed to distinguish between these categories. This section demonstrates the need for a nuanced approach to the syntax and morphology of Portuguese past participles, even considering only regular verbs. Section 4.3 sets out the phenomenon of double past participles, first from a cross-linguistic perspective and then focusing on Portuguese, before reviewing some previous approaches to the analysis of Portuguese double past participles in the literature.

Section 4.4 sets out my analysis of double past participles in Portuguese. I consider what their existence reveals about the interaction of morphology and syntax (section 4.4.1) before looking at how the complexities of category membership can be integrated into the analysis (section 4.4.2) and presenting my analysis of the position of past participles in the verbal paradigm (section 4.4.3). Section 4.4.4 deals with the lexical specification which is a crucial part of the double past participle phenomenon. The PFM rules required to provide correctly inflected past participle forms are presented in section 4.4.5 with worked examples in section 4.4.6. Section 4.4.7 provides rules for number and gender agreement, section 4.4.8 includes an analysis of the morphological syncretism of past participles in this model. In section 4.4.9 I present a formal approach to the morphology of past participles in periphrases, which is supplemented by an account of the syntactic and semantic properties of the passive voice alternation in section 4.4.10. Section 4.5 brings together conclusions from the chapter.

The analysis laid out in this chapter provides a foundation for later discussion of the diachronic development of distinct past participle forms, and the modelling of this development (Chapter 5).

4.1 Past participles and their properties

4.1.1 Functions of past participles in Portuguese

In Romance linguistics, past participles are standardly recognised as non-finite verb forms (Maiden 2016a, 507). Like other non-finite forms (infinitives and gerunds) in Portuguese, they do not express tense or mood (Cunha and Cintra 1990, 480). They can occur in several distinct syntactic contexts. The various functions of past participles in Portuguese are illustrated below, with the past participle underlined for clarity. Examples are adapted from an online manual of Portuguese

syntax (Bick 2001)[38] [my gloss and translation] and include verbs that are not part of the double past participle phenomenon; that is to say, the same form is used in all contexts. Examples (78) and (79) show past participles as analogous to adjectives, in their function of modifying a noun. Examples (81) to (85) show past participles in periphrases, which will be the main focus of this chapter.

(78) Attributive past participle

carros usados / velhos

cars.MPL used.PTCP.MPL / old.MPL

'used/old cars'

(79) Attributive past participle

uma casa muito aconchegada[39]/confortável

a.FSG house.FSG very comfortable.PTCP.FSG /comfortable.SG

'a very comfortable house'

(80) Attributive past participle (with dependents/arguments)

um ator geralmente muito apreciado pelo

an.MSG actor.MSG generally very appreciated.PTCP.MSG by.the.MSG

povo

people.MSG

'an actor generally very popular among the people'

---

[38] Bick's grammar is both a teaching manual of Portuguese syntax, and a manual for some Portuguese interactive grammar tools. The grammar is not specific to a particular variety of Portuguese (Brazilian or European).

[39] This form is related to the verb *aconchegar* lit. 'bring near', also 'make comfortable'.

(81) In active perfect verbal periphrasis with auxiliary verb *ter* 'have'

Ultimamente	tem	trabalhado	dia	e	noite
recently	have.PRES.3SG	worked.PTCP	day	and	night

'She/he has worked day and night recently.'[40]

(82) In passive periphrasis: 'action passive'[41] with auxiliary verb *ser* 'be'

Uma	hipopótama	branca	foi	(*nada)[42]
a.FSG	hippopotamus.FSG	white.FSG	was	(not at all)

vista	(por	um	turista)
seen.PTCP.FSG	(by	a.MSG	tourist.MSG)

'A white hippopotamus was (*not at all) seen (by a tourist)'

(83) In passive periphrasis: 'action passive' with auxiliary verb *ser* 'be'

A	grande	porta	foi	(*nada)	fechada
the.FSG	large.SG	door.FSG	was	(not at all)	closed.PTCP.FSG

(por	um	turista)
(by	a.MSG	tourist.MSG)

'The large door was (*not at all) closed (by a tourist)'

[40] This tense has a frequentative reading in Portuguese, so another translation would be 'She/he has been working day and night recently'.

[41] This corresponds to Embick's term 'eventive passive' (2004).

[42] The modifier *nada* is typically used with adjectives: see discussion in section 4.2.2.

(84) In passive periphrasis:[43] 'state passive'[44] with auxiliary verb *estar* 'be'

A hipopótama branca estava (nada)

the.FSG hippopotamus.FSG white.FSG was (not at all)

<u>cansada</u> (*por um turista)

tired.PTCP.FSG (by a.MSG tourist.MSG)

'The white hippopotamus was (not at all) tired (*by a tourist).

(85) In passive periphrasis: 'state passive' with auxiliary verb *estar* 'be'

A grande porta estava (nada) <u>fechada</u>

the.FSG large.SG door.FSG was (not at all) closed.PTCP.FSG

(*por um turista)

(by a.MSG tourist.MSG)

'The large door was (not at all) closed (*by a tourist)'

Unlike French or Italian, Portuguese does not have the possibility of using an auxiliary *be* in active perfective periphrases, that is, there are no complexities of auxiliary selection (McFadden 2007). The auxiliary *ter* 'have' is always used in active perfective periphrases, and the auxiliary *ser* 'be' or *estar* 'be' in passive periphrases. The term 'active periphrasis' therefore means 'active periphrasis with auxiliary *ter* 'have", and 'passive periphrasis' means 'passive periphrasis with auxiliary *ser* or *estar* 'be".

---

[43] The status of this construction as a passive can be disputed. It is essentially indistinguishable from a combination of the copula *estar* 'to be' plus an adjective. As I show in section 4.4.2.2 below, the past participle form in this construction can be more accurately considered a 'transposition' to adjective, which is still linked to its base lexeme (the verb).

[44] This corresponds to Embick's (2004) term 'stative passive'.

The past participle is also used in the participial absolute construction (also known as 'absolute participle construction' or *ablativus absolutus*): this occurs with low frequency.

(86) Participial absolute construction with transitive verb

<u>arrancada</u>	a	chave	da	vítima
wrenched.PTCP.FSG	the.FSG	key.FSG	of.the.FSG	victim.FSG
sumiu	na	mata		
fled.3SG	in.the.FSG	jungle.FSG		

'the key of the victim having been wrenched, he/she fled into the jungle'

(= Arrancou a chave da vítima,
wrench.3SG the.FSG key.FSG of.the.FSG victim.FSG
e sumiu na mata.'
and fled.3SG in.the.FSG jungle.FSG
'He/she wrenched the key from the victim, and fled into the jungle.')

The participial absolute clause has its own subject, and there is no agreement with the subject of the main clause. With a transitive verb (*arrancar* 'wrench'), the patient subject of the participle ('*a chave*') is the direct object of the corresponding active sentence (86). With an intransitive verb (*sumir* 'flee'), the subject of the participle is the subject of the corresponding active sentence (87).

(87) Participial absolute construction with intransitive verb

<u>sumido</u>	o	bandido	na	mata,	
vanish.MSG	the.MSG	bandit.MSG	in.the.FSG	jungle.FSG	
as	vítimas	se	consolaram		
the.FPL	victims.FPL	REFL	consoled.3PL		

'the victim having vanished into the jungle, the victims consoled each other'

(= O	bandido	sumiu	na	mata,	
the.MSG	bandit.MSG	fled.3SG	in.the.FSG	jungle.FSG	
e	as	vítimas	se	consolaram.	
and	the.FPL	victims.FPL	REFL	consoled.3PL	

'The bandit fled into the jungle, and the victims consoled each other)

(88) Further example of the participial absolute with transitive verb (Silva 2001, 87):

<u>Lida</u>	a	carta	o	poeta	se	suicidou.
read.PTCP.FSG	the.FSG	letter.FSG	the.MSG	poet.MSG	REFL	suicide

'The letter having been read, the poet committed suicide.'

In nearly all of the contexts in which past participles are found (78)-(88), the voice of the past participle is passive. For example, *carros usados* 'used cars' in (78) could be paraphrased as 'cars that have been used', not '*cars that have used'. The exceptions are (81), where the past participle is in an active periphrasis (*tem trabalhado* 'she has worked' rather than '*she has been worked'), and (87), which includes the past participle of an intransitive verb.[45] Excluding the participial absolute, a

[45] In this context, past participles of intransitive verbs are typically active, whereas past participles of transitive verbs are passive (Cunha and Cintra 1990, 492).

somewhat marginal context, we can generalise that past participles are active if they appear in an active periphrasis with the auxiliary verb *ter* 'have', and are otherwise passive.

In the double past participle phenomenon, which is the focus of this chapter, these functions are divided up between more than one morphological form (section 4.2). In the next section, I discuss what these morphological forms may look like in Portuguese.

4.1.2 Forms of past participles in Portuguese

Regular Portuguese verbs have a single inflected past participle that carries out all of the functions exemplified in (78) to (88). This regular past participle is formed of the basic verb stem plus *-ado* (first conjugation) or *-ido* (second and third conjugations).[46] This is the case, for example, for the verbs *cantar* 'sing', *comer* 'eat' and *partir* 'break', which represent the three major conjugation classes.

Some Portuguese verbs have an irregular past participle form. For some verbs, like *fazer* 'do', past participle *feito*, this is just one irregularity in a verb paradigm that has many irregular forms. For other verbs, like *escrever* 'write', past participle *escrito*, the past participle is the only irregular form in the verb paradigm. Table 9 summarises this information.

Verb type	**Infinitive**	**Meaning**	**Past participle**
Regular verb: first conjugation	CANTAR	sing	*cantado*
Regular verb: second conjugation	COMER	eat	*comido*
Regular verb: third conjugation	PARTIR	break	*partido*
Irregular verb (many irregularities)	FAZER	do	*feito*
Irregular verb (only past participle is irregular)	ESCREVER	write	*escrito*

Table 9 Portuguese past participles by verb type

[46] Some past participles also show agreement in number and gender: see section 4.2.2.1.

To show the full range of possible morphological forms for the past participle, Table 10 lists a set of verbs which have an irregular past participle form (some of which also have a regular past participle form: see section 4.3). The stem, which may be irregular, is represented by 'X' in this table.

Verb	**Meaning**	**Conjugation**	**Irregular PTCP**	**Pattern**
ABRIR	open	3	*aberto*	X t o
ACEITAR	accept	1	*aceite*	X e
ACENDER	light/turn on	2	*aceso*	X s o
COBRIR	cover	3	*coberto*	X t o
DIZER	say	2	*dito*	X t o
ELEGER	choose/elect	2	*eleito*	X t o
ENTREGAR	hand over	1	*entregue*	X e
ESCREVER	write	2	*escrito*	X t o
EXPRIMIR	express	3	*expresso*	X s o
FAZER	do	2	*feito*	X t o
GANHAR	earn	1	*ganho*	X o
GASTAR	spend	1	*gasto*	X o
LIMPAR	clean	1	*limpo*	X o
MORRER	die	2	*morto*	X o
PAGAR	pay	1	*pago*	X o
PÔR	put	2	*posto*	X t o
PRENDER	arrest	2	*preso*	X s o
SUSPENDER	suspend	2	*suspenso*	X s o
VER	see	2	*visto*	X t o

Table 10 Morphological patterns of irregular past participles

As the table shows, the morphological patterns found in irregular past participle forms are quite diverse, and are also quite distinct from the patterns of regular past participle forms. The proper analysis of these morphological patterns in my theoretical framework is taken up in section 4.4.5.

Having seen the basic facts of the functions and forms of past participles in Portuguese, I now move on to discuss one of their complex characteristics: their lexical category properties.

4.2 Lexical category of past participles

One recurring theme of previous work on past participles is whether they are more verbal, or more adjectival. For example, Embick (2004, 355) refers to the 'familiar' division of past participles into adjectival and verbal types, and Laurent (1999, 356) also discusses this distinction. This question merits some consideration here due to its implications for a theoretical analysis of Portuguese past participles. At a simple level, a standard LFG or PFM approach presupposes that lexical items belong to a lexical category, and past participles must therefore be classified in some way. In this section, I discuss standard approaches to the lexical categories of verbs and adjectives (section 4.2.1) before considering Portuguese past participles according to three types of criteria (section 4.2.2).

4.2.1 Differentiation of verbs and adjectives

Lexical categories present various complexities, resulting in a great range of treatments, as discussed for example in Baker and Croft's (2017) survey, which covers approaches to lexical categories from both 'functionalist' and 'formalist' perspectives. One reason for the range of treatments of lexical category is that several properties can be considered relevant, including the distribution, function and morphological form of a word.

According to Baker (2003), 'the most traditional and widespread view about category distinctions is that they are essentially morphological in nature'. Baker traces this view to ancient Greek and Roman grammarians who considered the inflectional patterns of words as the definitive guide to

what we now think of as lexical category. In this view, the fact that a given word shows particular morphological patterns marks it out as a verb, noun or adjective. But category can also be defined by completely opposing characteristics: in Distributed Morphology (Halle and Marantz 1993; 1994), 'roots', which are basic lexical items, have no category, but receive a category when they are inserted in a particular syntactic context. This is a fully syntactic view of lexical categories.

Apart from these two extremes of 'purely morphological' and 'purely syntactic' views of lexical category, there are many other possibilities. A familiar perspective from formal theoretical syntax is Chomsky's (1970) analysis of categories using two cross-cutting binary features (±N and ±V) to define nouns (+N, -V), verbs (-N, +V), adjectives (+N, +V) and adpositions (-N, -V). This approach has been subject to criticism; for example Baker (2010) states it is 'widely recognised to have almost no content in practice'. Baker himself favours a stricter view of categories, to permit effective typological comparison across languages. In his proposals, each of the three lexical categories has a single defining property: nouns have a referential index, verbs have a specifier, and adjectives are neither nouns nor verbs. Other classes of words, such as adpositions, are functional rather than lexical categories. This approach has the advantage of giving a simple definition for each of the three categories, but the negative definition for adjectives is clearly of limited use in distinguishing adjectives from verbs, and the analysis as a whole presupposes a particular theoretical foundation which does not easily transfer to different theoretical contexts.

A typological approach to the definition of lexical categories is given by Dixon (2004). In this view, the word classes of a language have both 'a prototypical conceptual basis' and 'prototypical grammatical function(s)' and can be recognised on the basis of 'internal grammatical criteria for that language', such as morphological processes and syntactic distribution (Dixon 2004, 2). Across languages, lexical categories can be identified by their 'common semantic core'. Dixon's focus is on adjectives. He considers 'adjectives in the narrow sense – that is, descriptive adjectives such as 'red', 'heavy', and 'loyal"; this class has a 'common semantic core' of words relating to dimension,

age, value and colour (Dixon 2004, 3–4). Dixon states that he is 'for the most part, concerned just with morphologically simple roots, not with derived stems', even while recognising that a class of adjectives in a language may contain forms derived from verbs.

This caveat is also reflected in Baker's treatment of participle forms, which is given in a footnote (Baker 2003, 198, footnote 2). He states that 'Participial forms of the verb can also modify nouns, as in *a shining light* […] I leave a detailed study of participles to future research.' It is clear that treatments of lexical categories, whether theory-specific or theory-neutral, tend to focus on the simpler cases of uninflected or prototypical simple forms. Participles are problematic because they are lexically related to verbs, but have adjectival properties. Far from being specific to Portuguese or to past participles, this difficulty of analysis is common to many types of lexically related forms, including non-finite verb forms, across languages. An explicit recognition of the variety of different forms with complex relationships to other lexical items, as presented by Spencer (2013) and introduced in section 3.3.2 above, is required for an effective analysis. This is discussed further in section 4.4.2.2.

It is necessary to consider the typical properties of adjectives and verbs, to be clear about the problems past participles present (section 4.2.2). Dixon's typical semantic (or 'conceptual') properties of adjectives and verbs are not useful for these purposes, as we are not looking at straightforward cases. We must therefore focus on grammatical properties. Some properties of verbs and adjectives seen across languages, as given by Dixon, are summarised in Table 11.

Verbs	**Adjectives**
Their primary function is as the head of a predicate	They may fill the copular complement slot
They can be the nucleus of a transitive or intransitive clause	They can be the nucleus of an intransitive clause in some languages
In European languages, their morphological categories typically include tense, aspect and mood	In European languages, they have a similar morphology to nouns
In some languages, they inflect for person and number of the subject	In European languages, they typically have the morphological property of gender
	In Latin, their morphological categories include number and case
	They tend to have two roles: 'In a statement that something has a certain property' or 'As a specification that helps focus on the referent of the head noun in an NP that relates to a predicate argument'

Table 11 Typical properties of verbs and adjectives

These major properties of verbs and adjectives are worth bearing in mind for considering the category properties of the past participle in different contexts (where it might be more verbal or adjectival). With these properties laid out, I now move on to a more fine-grained approach to grammatical categories, which considers three sets of criteria: internal syntax, distribution, and morphosyntax. These criteria can productively be applied to inflected or derived forms, thus avoiding some of the problems of other approaches to lexical category.

4.2.2 Distinguishing the lexical category of Portuguese past participles

In this section, I investigate evidence for the lexical category of Portuguese past participles, showing that, like many other non-finite forms across languages, they are not straightforwardly of a single lexical category. I assess Portuguese past participles in their different functions according to criteria laid out by Lowe (2019) as 'distributionalist criteria for distinguishing categories'. These are a more explicit example of the types of criteria that Dixon calls 'internal grammatical criteria' for each language.

a. Internal syntax: the internal structure of the phrase, for example whether it contains determiners, adjectives, objects, adverbs.
b. Distribution: the distribution of the phrase at a clausal level, for example whether a phrase can appear in the same structural positions, and fill the same grammatical functions, as unambiguous noun phrases, or verb phrases, etc.
c. Morphosyntax: the morphosyntactic properties of the head of the phrase, for example whether it shows the agreement features typical of a verb, an adjective, etc.

(Lowe 2019, 2)

By 'distributionalist' criteria, Lowe means that these are broadly syntactic in nature, as opposed to 'essentialist' criteria, which involve non-syntactic aspects such as lexical semantics in the definitions of category.

In straightforward cases, all three criteria point towards the same lexical category. However, this is not always the case. Lowe proposes theoretical categories, discussed further in section 4.4.2.1, of 'mixed categories', where the internal syntax is inconsistent, and 'mismatched categories', where evidence from internal syntax contradicts evidence from distribution or morphosyntax.

Table 12 shows properties of verbs and adjectives in Portuguese across the three types of criteria: morphosyntax, distribution and internal syntax.

	Verbs	**Adjectives**
Morphosyntax	Agree in person and number with the subject Inflect for tense, mood and aspect	Agree in number and gender with the noun they modify
Distribution	Predicate of a main clause: *Ela lavou o carro* 'she washed the car'	Attributive use: *carros velhos* 'old cars' Predicative use: *o carro é velho* 'the car is old'
Internal syntax	Take arguments such as objects (transitive verbs) Can be modified by adverbs Can be negated	Cannot have arguments Can be modified by adjective-specific intensifiers such as *nada*[47]

Table 12 Lexical category properties of verbs and adjectives in Portuguese

Portuguese past participles are not easily identifiable as belonging to a single lexical category, as I demonstrate in the following sections. Morphologically, they do not agree in person, or inflect for tense, mood or aspect, but they sometimes agree in number and gender. Their internal syntax and distribution are sometimes hard to assess since past participles can occur within verbal periphrases, as well as independently within a clause. The next sections review the evidence in more depth.

4.2.2.1 Morphosyntax of past participles

First, I consider morphosyntax, which is the most relevant facet for my PFM analysis. There are two distinct patterns of behaviour. In the active perfective periphrasis shown in (89), (90) and (91),

[47] In testing this property it is important to distinguish between adjective-specific intensifiers and general adverbs. Bick (2001, 69) suggests *nada* can be used for this test.

the past participle does not inflect for person or number, appearing in a default form which consistently ends *-o*. Although the periphrasis as a whole has a tense and mood value, this is not attributable to the past participle, which is compatible with various tenses and moods, as shown in the examples.

(89) Terá comido
have.3SG.FUT eaten.PTCP
'She/he/it will have eaten'

(90) Tinha comido
have.1/3SG.IPFV eaten.PTCP
'I/she/he/it had eaten'

(91) Se tivesse comido, …
If have.1/3SG.IPFVSUBJ eaten.PTCP
'If I/he/she/it had eaten, …'

In the passive periphrastic context, the past participle agrees in number (94) and gender (95) with the subject, employing the default agreement patterns for Portuguese adjectives, as demonstrated in (92) and (93) for number and gender respectively. These default number and gender markers are shown in Table 13. The full system of number and gender agreement in Portuguese is more complex but the details are not relevant here.

	Masculine	**Feminine**
Singular	*-o*	*-a*
Plural	*-os*	*-as*

Table 13 Number and gender agreement in Portuguese

(92) Os bolos eram brancos
the.MPL cakes.MPL be.3PL.IPFV white.MPL
'The cakes were white'

(93) A bolacha era branca
the.FSG biscuit.FSG be.3SG.IPFV white.FSG
'The biscuit was white'

(94) Os bolos foram comidos
the.MPL cakes.MPL be.3PL.PAST eaten.PTCP.MPL
'The cakes were eaten'

(95) A bolacha foi comida
the.FSG biscuit.FSG be.3SG.PAST eaten.PTCP.FSG
'The biscuit was eaten'

As these examples show, the past participle in this passive context agrees in number and gender like a simple adjective, whereas in the active context, there is no agreement.

4.2.2.2 Distribution of past participle

Next, I consider the past participles' properties in terms of distribution, looking at the possible structural positions of the past participle.

Past participles may have the distribution of attributive adjectives, as shown in examples (78) and (79), repeated here as (96) and (97). The words *velhos* 'old' and *confortável* 'comfortable', which can replace the past participles in these sentences, are canonical adjectives.

(96) Attributive past participle
carros usados / velhos
cars.MPL used.PTCP.MPL / old.MPL
'used/old cars'

(97) Attributive past participle

uma	casa	muito	aconchegada /confortável
a.FSG	house.FSG	very	comfortable.PTCP.FSG /comfortable.SG

'a very comfortable house'

Past participles cannot be the predicate of a main clause, although they can occur in active perfective verbal periphrases with an auxiliary *ter* 'have', in passive periphrases with the auxiliaries *ser* or *estar* 'be', and in the participial absolute. Any classification of past participles within these distributions as either verbal or adjectival would be circular, since only past participles occur in these contexts.

The evidence from distribution is therefore also complex; past participles can occur in a variety of structural positions, rather than having the same distribution as verbs or adjectives.

4.2.2.3 Internal syntax of past participles

The evidence from internal syntax is also somewhat complex. In different contexts, past participles may be compatible with either the modifiers and arguments typical of adjectives, or those typical of verbs.

Examples (83) and (85), repeated here and renumbered as (98) and (99), demonstrate this. The past participle in the action passive (98) is incompatible with the modifier *nada*, which is typically used with adjectives (Bick 2001, 69). Example (98) also shows that in this context the past participle can be used with an agent by-phrase, here *por um turista* 'by a tourist'. An agent by-phrase would typically be considered an argument (or adjunct) of a verb phrase. However, there is an element of circularity here, because agent by-phrases occur with passives, which always involve a past participle, so this test is not particularly revealing. In the state passive (99) the past participle is compatible with the modifier *nada* lit. 'nothing', here 'at all', but is not compatible with the agent by-phrase. These two examples show that the past participle form may take the modifiers and arguments typical of either verbs or adjectives.

(98) In passive periphrasis: 'action passive' with auxiliary verb *ser* 'be' (more verb-like)

A	grande	porta	foi	(*nada)	fechada
the.FSG	large.SG	door.FSG	was	(not at all)	closed.PTCP.FSG

(por um turista)

(by a.MSG tourist.MSG)

'The large door was (*not at all) closed (by a tourist)'

(99) In passive periphrasis: 'state passive' with auxiliary verb *estar* 'be' (more adjective-like)

A	grande	porta	estava	(nada)	fechada
the.FSG	large.SG	door.FSG	was	(not at all)	closed.PTCP.FSG

(*por um turista)

(by a.MSG tourist.MSG)

'The large door was (not at all) closed (*by a tourist)'

The internal syntax of the past participle in attributive contexts can be analysed as more adjectival when it has no dependents, but more verbal when it does have dependents (Bick 2001, 67). The internal syntax of the past participle in the participial absolute is far from clear, as it cannot be modified or negated like a verb, but has arguments like a verb. The internal syntax of the whole active periphrasis is clearly verbal, and so we might also say that the past participle in this context has verbal internal syntax.

The internal syntax of the past participle is therefore not straightforward.

4.2.2.4 Summary of lexical category properties of past participles

Evidence for the category properties of past participles, brought together from the preceding sections, is summed up in Table 14.

Function	**Example**	**Morphosyntax**	**Distribution**	**Internal syntax**
Attributive past participle	(78), (79)	Adjectival: agrees in number and gender	Adjectival: can be replaced by an adjective	Adjectival
Attributive past participle with dependents or arguments	(80)	Adjectival: agrees in number and gender	Specific to the past participle	Verbal
Past participle in active verbal periphrasis	(81)	Verbal: no number or gender agreement	Specific to the past participle	Verbal
Past participle in action passive periphrasis with auxiliary *ser*	(82), (83)	Adjectival: agrees in number and gender	Specific to the past participle	Verbal
Past participle in state passive with auxiliary *estar*	(84), (85)	Adjectival: agrees in number and gender	Adjectival: can be replaced by an adjective	Adjectival
Participial absolute	(86), (87)	Adjectival: agrees in number and gender	Specific to the past participle	Mixed

Table 14 Category properties of past participles

As Table 14 shows, taking into account the three different sets of properties of morphosyntax, distribution, and internal syntax, past participles can have a whole range of behaviour, from being purely verbal, to being purely adjectival, and various combinations of properties in between. It is

clear that 'past participles' in Portuguese, considered as a whole class, are not straightforwardly either verbs or adjectives, like other derived or non-finite forms in many languages. An effective theoretical approach must be able to capture this. My analysis, presented in section 4.4.2, focuses on the morphological aspects of the past participle and therefore morphosyntactic criteria are weighted more heavily. This analysis builds on the work of Spencer (2013) which develops the distinction between simple inflected forms and forms related to their base lexeme through some other type of lexical relatedness.

In general in my analysis, the past participle is generated as a member of the verbal paradigm (and the morphological category V) in PFM rule blocks, although in LFG c-structure terms it could sometimes be treated as category Adj and sometimes as part of a periphrasis with the category V. This allows the c-structure analysis to reflect the different syntactic distribution and behaviour of the past participle (aside from its morphological structure), reflecting the LFG treatment of different aspects of the grammar on their own terms.

In the next section, I introduce the double past participle phenomenon, which is the focus of this chapter.

4.3 The double past participle phenomenon

The vast majority of verbs across the Romance languages have only one past participle form, which is used for all the functions laid out in section 4.1.1. However, across the Romance family, there are a number of varieties in which some verbs have more than one past participle form. My focus is on those Portuguese verbs with two past participle forms which carry out different functions.

A simple constructed example is given here to introduce the phenomenon. This example is discussed further in section 4.3.2.

(100) A Maria tinha acendido as luzes.
the Maria had turned-on.REGPTCP the lights
'Maria had turned on the lights.'

(101) As luzes foram acesas pela Maria.
the.FPL lights.FPL were turned-on.IRREGPTCP.FPL by.the Maria
'The lights were turned on by Maria.'

The Portuguese double past participle phenomenon is particularly notable because it throws into contrast an interesting distinction between active periphrastic and passive contexts, which is not reflected in other Romance languages that have two past participle forms.[48] In these languages, the typical split is between contexts where the past participle has the internal syntax of an adjective and contexts where its internal syntax is verbal, meaning the morphological distinction aligns (at least under some views) with a lexical category distinction. As I discuss in 4.3.3, the Portuguese split can be viewed as an exception to the usual syncretism between the past participle used in active and that used in passive contexts, which makes it distinct from similar phenomena in other Romance varieties (Maiden 2016b; 2013).

The double past participle phenomenon can be treated as a kind of mismatch between the expected alignment of forms and functions: from a cross-Romance perspective, we expect one form to correspond with both active and passive functions, whereas in these particular Portuguese verbs, this expectation is not met (different forms are used for each function). Seen from the other perspective, though, the double past participle phenomenon can be identified as a more straightforward alignment, with the usual cross-Romance pattern actually showing a mismatch between number of forms and number of functions. Whichever perspective we take, the double

[48] A possible exception is some Galician verbs: see section 4.3.1.

past participle phenomenon is clearly worth studying to shed further light on morphosyntactic mismatches.

This section therefore introduces double past participles from a cross-linguistic perspective (section 4.3.1), before focusing on the phenomenon in Portuguese (section 4.3.2), and discussing previous approaches to the Portuguese data (section 4.3.3).

4.3.1 Double past participles from a cross-linguistic perspective

In this section, I briefly consider some of the languages in which double past participles exist, both within and outside the Romance family. I mention similarities and differences between the Portuguese double past participle phenomenon, and the phenomenon seen in other languages.

A number of verbs in Spanish and Galician have two forms. In Spanish, the forms are argued (see for example Gibert Sotelo 2017) to be split in distribution between so-called adjectival contexts ('attributive past participle' and 'state passives' with auxiliary *estar* 'be'), where the irregular form e.g. *lleno* 'full' is used, and verbal contexts (active periphrastic and 'action passives' with auxiliary *ser* 'be'), where the regular form e.g. *llenado* 'filled' is used.

An example of the Spanish split between verbal and adjectival contexts, given by Gibert Sotelo (2017, 237), is:

(102)	El	vaso	está	<u>lleno</u>		'Perfective adjective'
	the	glass	be.3SG.PRES	fill.PTCP		[= 'state passive']
	'The glass is full'					
(103)	El	vaso	ha	sido	<u>llenado</u>	'Verbal past participle'
	the	glass	have.3SG.PRES	be.PTCP	fill.PTCP	[= 'action passive']
	'The glass has been filled'					

In Galician, there are also some verbs with an adjectival/verbal distinction along the lines seen in Spanish (Alvarez Blanco, Monteagudo Romero, and Regueira 1986, 401–2), while other Galician

verbs may behave more like the Portuguese verbs, with a split between the active verbal context and passive contexts (Thomas 2018, 103–7). Further research is needed to clarify exactly what the patterns of usage are; unlike Spanish or Portuguese past participles, the Galician past participles have been the subject of little linguistic research.

Multiple past participle forms have also been attested in some dialects of Italy. I refer interested readers to Ledgeway (2016) and references therein for Southern Italian (Salentino and Calabrian), Da Tos (2012) for Venetan, and Thornton (2011) for some 'overabundant' past participle forms in Standard Italian. In some of these varieties, the different forms are not divided along syntactic lines, instead being split by sociolinguistic distinctions. In others, there is a division like in Spanish between adjectival and verbal contexts.

Beyond the Romance languages, some other Indo-European languages show comparable patterns, although these are often substantially different in nature. For example, all Russian verbs systematically have a 'long' and a 'short' past participle form, which have a difference in interpretation: long forms are purely stative, while short forms can have an eventive or stative interpretation (Borik 2014). In the Germanic languages, it has been argued that, for at least some speakers, some English verbs have competing past participle forms:

(104) John has <u>shaved</u> his head.

(105) John's head was <u>shaven</u>.

The 'shaved' form is preferred when a verbal reading is coerced by the addition of an agent by-phrase:

(106) *John's head was <u>shaven</u> by Peter.

(107) John's head was <u>shaved</u> by Peter.

The English data are discussed by Embick (2004) who distinguishes between 'stative', 'resultative' and 'eventive' contexts. Where a distinction in forms occurs in English, stative forms on one side (108) differ from resultative and eventive forms on the other (109):

(108) John's head was shaven. (Stative)

(109) John's head was shaved (by Peter). (Resultative or eventive)

Embick (2004, 360–61) claims that both 'stative' and 'resultative' functions are traditionally classified as 'adjectival' uses of the past participle, and concedes that 'given that the forms are often the same, the differences between stative and resultative in these environments [appearing in both attributive and predicative position] can be quite subtle'.

There are clear difficulties in distinguishing these different classifications of past participles in English, and the tests used by Embick do not necessarily transfer effectively to Portuguese or the other Romance languages. In my discussion of the distribution of the two different past participle forms, I therefore focus on the different functions of the past participle in Portuguese, as set out above in section 4.1.1.

4.3.2 Double past participles in Portuguese

There is considerable variation across Portuguese in the existence of two past participle forms and their usage (Schwenter et al. 2019). The split that I am focusing on involves a distinction between active periphrastic and passive functions. That is to say, one form (*regular*) is used in active periphrases, which are formed with the auxiliary verb *ter* 'have'. The other form (*irregular*) is used

in passive contexts, both action passives with auxiliary *ser* 'be' and state passives with auxiliar *estar* 'be', as well as in attributive contexts.[49]

There is substantial variation in the terminology that has been used to refer to the two distinct forms of the past participle, for verbs that have two forms. Most of this terminology is based on the morphophonology of the forms. One form is characterised as 'irregular', 'short', 'athematic' (lacking a thematic vowel), 'rhizotonic' (root-stressed), or 'truncated'. The other form is characterised as 'regular', 'long', 'thematic', or 'arrhizotonic'. Table 15 sets out terminology used in different contexts, separating the forms according to their function in European Portuguese.

---

[49] Searches on the *Corpus do Português* (Web/Dialects) show that the irregular form is used in the participial absolute construction in European Portuguese. This aligns with evidence from Brazilian Portuguese:

Mortos	os	soldados	inimigos,	
kill.PTCP-S.MPL	the.MPL	soldiers.MPL	enemy.MPL	
a	batalha	foi	considerada	vencida
the.FSG	battle.FSG	was	considered.PTCP.FSG	won.FSG

'The enemy soldiers having died, the battle was considered won.'

(Silva 2001, 7–8)

Here the irregular past participle form *morto* of the verb *matar* 'kill' is used in the participial absolute construction. The form agrees in gender and number. Fong (2019) presents unattributed data for Brazilian Portuguese claiming that the regular form is used in this construction, but I disregard this here.

Used in passive periphrases in EP	**Used in active periphrases in EP**	**Terms used by:**
e.g. *aceso* 'lit', *entregue* 'handed over'	e.g. *acendido* 'lit', *entregado* 'handed over'	
Irregular	Regular	e.g. Schwenter et al. (2019), Lobato (1999)
Short	Long	e.g. Nevins & Rodrigues (2014)
Athematic (has no theme vowel)	Thematic (has a theme vowel)	e.g. Chagas de Souza (2011)
Rhizotonic (root-stressed)	Arrhizotonic (not root-stressed)	e.g. Maiden (2016b) on Latin
Strong	Weak	e.g. Ledgeway (2016) on Italian
Truncated		e.g. Svensen (2013) on Spanish ('*truncos*') [50]

Table 15 Terminology for past participle forms

I choose to use the terms *regular* and *irregular*. In the following discussion I also sometimes refer to *active type* and *passive type* past participles, particularly when dealing with verbs that do not have

[50] A description of Galician grammar (Alvarez Blanco, Monteagudo Romero, and Regueira 1986, 401–2) distinguishes between Truncated irregular past participles (first conjugation or *-ar* verbs), formed with regular stem + vowel, and Strong irregular past participles (second and third conjugation, *-er* and *-ir* verbs), formed with irregular stem + vowel.

distinct forms for these two functions, and therefore have no regular versus irregular distinction. When relevant, I gloss regular and irregular past participles of double past participle verbs as REGPTCP and IRREGPTCP.

The set of verbs that exemplify the double past participle pattern is not fixed, and differs between Brazilian and European Portuguese, as well as between individual speakers. Further variation may also exist at a smaller regional scale, but this topic has not been studied in detail. Here my focus is on European Portuguese: the use of irregular past participles alongside regular past participles is claimed to be more frequent in this variety than in Brazilian Portuguese (Schwenter et al. 2019, 75). In addition, there are various complexities specific to Brazilian Portuguese which complicate the phenomenon. For example, Brazilian Portuguese has a set of 'innovative' irregular past participles which may be preferentially used in active contexts (Scher, Lunguinho, and Rodero-Takahira 2013), contrary to the normal use of irregular past participles in European Portuguese (see section 4.3.3). While considering the Brazilian Portuguese data and previous analyses of the data adds vital context to the discussion, my focus in this chapter is on European Portuguese, and I do not present a full analysis of Brazilian Portuguese here, reserving this for Chapter 5.

Another important fact about the double past participle phenomenon is that it is not diachronically stable, although there is evidence for verbs with two past participle forms throughout the history of Portuguese. Over time, verbs may gain or lose past participle forms. This is discussed in Chapter 5, where I adapt the formal theoretical analysis presented in this chapter to account for the diachronic facts.

The extent to which the double past participle phenomenon is still extant in European Portuguese in its true sense (i.e. one form used for active contexts and a different form for passive contexts) has been debated. For example, Bonami and Luís (2015) give evidence from a contemporary corpus of European Portuguese journalism (CETEMPúblico) that both regular and irregular forms of the verb *encarregar* 'entrust/charge someone with something' are used in both active and passive

contexts, suggesting the forms are in free variation rather than distinguished by grammatical properties. These are interesting datapoints which undoubtedly reflect the high levels of variation in usage associated with the regular and irregular past participle forms, as discussed below (section 4.3.3). However, such evidence does not rule out the continued existence of the double past participle phenomenon for some speakers and/or for some verbs, and we therefore need an effective grammatical account of the phenomenon.

One verb for which there is good evidence that the two past participle forms still exist and are distinguished along active/passive lines in European Portuguese is *acender*, 'light/turn on'. This was confirmed in a grammaticality judgement study with 43 European Portuguese speakers which I undertook in Lisbon in 2017.[51] Speakers gave the regular form of *acender* an average rating of 3.55/4 in the active context, and 1.9/4 in the state passive context, while the irregular form of *acender* was given an average rating of 2.57/4 in the active context, and 3.9/4 in the state passive context (Thomas 2018), demonstrating that speakers prefer the regular form in the active context and the irregular form in the state passive context for this verb.[52]

(110) A Maria tinha <u>acendido</u> as luzes.
the Maria had turned-on.REGPTCP the lights
'Maria had turned on the lights.'

[51] The research received ethical approval from the University of Oxford's Central University Research Ethics Committee, reference R51936/RE001.

[52] The differences in ratings of the regular and irregular forms of *acender* in these contexts are statistically significant ($p<0.001$).

(111) As luzes foram acesas pela Maria.
the.FPL lights.FPL were turned-on.IRREGPTCP.FPL by.the Maria
'The lights were turned on by Maria.'

Example (110) illustrates the active periphrastic context. The regular past participle form *acendido* appears here, and does not show any number or gender agreement. Example (111) illustrates the passive context. The irregular past participle form *aceso* appears here, and it agrees in gender and number with the subject (feminine plural), giving the form *acesas*.

There are therefore two differences in the past participle used in each example: the regular versus irregular form, and the agreement properties.

To compare this with a regular Portuguese verb, which uses the same past participle form in all contexts, consider the constructed examples (112) and (113).

(112) A Joana tinha comido os bolos.
the Joana had eaten.PTCP the cakes
'Joana had eaten the cakes.'

(113) Os bolos foram comidos pela Joana.
the cakes were eaten.PTCP.MPL by.the Joana
'The cakes were eaten by Joana.'

The verb *comer* 'eat' has only one past participle form, *comido*, rather than having both a regular and an irregular form. This form appears in both the active and passive sentences, although in the passive example (113) it agrees in number and gender with the subject, with agreement morphology *-os*.

Portuguese double past participles have received a considerable amount of attention, particularly at the descriptive level, and more recently from certain theoretical perspectives too (e.g. Nevins and Rodrigues 2014). The next section surveys the major themes of this work.

4.3.3 Previous approaches to Portuguese double past participles

In traditional grammars, such as Cunha (1972), and pedagogical material such as Lopes (1987), double past participles are generally presented as a regular and well-understood phenomenon. A typical presentation includes a list of verbs, with their two forms, and an illustration of the contexts in which the irregular form and the regular form are generally used. This type of description tends to oversimplify the phenomenon, as well as normally including many verbs that no longer have two past participle forms. Some of the forms that are claimed to exist as past participles are purely adjectival in synchronic terms. These forms, which have origins as past participles, can come to have semantics distinct from their base lexical form. For example, the Portuguese form *tinto* used to function as an irregular past participle form of *tingir* 'dye, stain' (114) but now only functions as an adjective ('red') or noun ('red wine') in the modern language (116), with *tingido* as the only possible past participle of the verb (117). Examples are from the *Corpus do Português* (Davies and Ferreira 2006; Davies 2016).

(114) O fato novo de algodão, tinto de lama
the suit new of cotton, stain.IRREGPTCP of mud
'His new cotton suit, stained with mud'
(*Corpus do Português: Genre/Historical,* 1910)

(115) o xale tinto muito cingido
the shawl dye.IRREGPTCP very tight
às omoplatas magras
to.the shoulder.blades thin
'her dyed shawl very tight around her thin shoulder blades'
(*Corpus do Português: Genre/Historical,* 1878)

(116) O vinho tinto dá um sabor especial ao molho
the wine red gives a flavour special to.the sauce
'The red wine gives a special flavour to the sauce'

(*Corpus do Português:* Web/Dialects)

(117) se seu cabelo já foi tingido
if your hair already was dyed.REGPTCP
ou tratado com amônia
or treated with ammonia
'If you hair has already been dyed or treated with ammonia'

(*Corpus do Português:* Web/Dialects)

Lists of double past participle forms often include this type of lexical item. The exact verbs included in these lists varies considerably between accounts, as can be seen by comparing the list in the pedagogical *Dicionário de Verbos* (J. A. Lopes 1987, 413–15) with Cunha and Cintra's (1990, 441–42) list.

By contrast, the complexities of the double past participle phenomenon have long been recognised by some linguists. In his *Dificuldades da Língua Portuguesa* ([1908] 1957), the grammarian Said Ali notes that there is no accurate general rule, and that idiosyncratic variation from verb to verb in the use of the two past participle forms has to be taken into account, a view that is now widely accepted.

Looking beyond traditional grammars, a more rigorous descriptive approach has been taken in recent studies of the phenomenon, identifying that rule-based approaches to double past participle forms do not describe usage accurately, because different verbs are used in different ways, in line with Said Ali's claims. For European Portuguese, Villalva and Almeida (2004) use an acceptability judgement questionnaire to assess speakers' perceptions of usage, concluding that there is

substantial variation in speakers' application of the prescriptive rules, and that verb-to-verb variation is also crucial. Móia (2004) draws similar conclusions based on a corpus study, noting that many lists of double past participle forms fail to distinguish between different patterns of behaviour, and incorporate forms which act only as adjectives, such as *tinto* 'red'.

As for Brazilian Portuguese, Miara (2013) compares use of the two past participle forms in a written corpus of Brazilian Portuguese from a regional newspaper, identifying that verb-to-verb variation (i.e. that the irregular past participle is simply more likely to be used for some verbs than for others) is the most relevant factor of variation, as well as finding distinctions between the active and passive. Sociolinguistic variables of age, gender and education were not relevant. Silva (2008) also investigates regular versus irregular participles in Brazilian Portuguese, finding that regular past participles are favoured by speakers with a lower level of education, and irregular ones occur more frequently in the speech of more educated speakers, with a high level of variation in most speakers' use of the two forms.

Most recently, Schwenter et al. (2019) use a multifactorial approach to study the relevance of various properties for modern usage of irregular versus regular past participle forms in the active context with 'have' auxiliaries, in European and Brazilian Portuguese. They find among other conclusions that lexical specification is the most important factor: that is to say, the irregular past participles of some verbs are more frequently used in this context than the irregular past participles of other verbs.[53] An effective theoretical account of the phenomenon therefore needs to capture

[53] Schwenter et al. (2019) do not consider the distribution of irregular vs. regular forms in other contexts, which reduces the usefulness of their paper in considering the resilience of the double past participle phenomenon. Verbs which show categorical use of the irregular form in the active

the fact that having two past participle forms is a lexical property of certain verbs, which may vary from speaker to speaker.

Recent work has also focused on Brazilian Portuguese innovative past participles, which can be characterised as new irregular forms. Chagas de Souza (2007, 126) gives the following examples [my gloss, original translation]:

Innovative (non-standard) past participles in the first conjugation:

(118) *chegar* 'arrive' > *chego* 'arrived'

A	carta	tinha	chegado/chego
the	letter	had	arrived.REGPTCP /IRREGPTCP

'The letter had arrived'

Innovative (non-standard) past participles in other conjugations:

(119) *trazer* 'bring' > *trago* 'brought' (second conjugation)

O	professor	tinha	trazido/trago	o	livro
the	teacher	had	brought.REGPTCP /IRREGPTCP	the	book

'The teacher had brought the book'

(120) *pedir* 'ask' > *peço* 'asked' (third conjugation)

O	presidente	tinha	pedido/peço	paciência
the	president	had	asked.for.REGPTCP /IRREGPTCP	patience

'The president had asked for patience'

---

context might no longer truly be 'double past participle' verbs, since this implies that the irregular form will be preferentially used in all contexts.

These innovative forms are not inherited from Latin or found in historical Portuguese data: they have been introduced to the language recently. The masculine singular of the innovative past participle is often identical in form to the first person singular present indicative form of the verb, e.g. *chego* 'arrived/I arrive' (innovative past participle, and first person singular present indicative). This is particularly obvious for verbs like *pedir* 'ask' (120). The innovative past participle *peço* 'asked' is identical to the irregular first person singular indicative form, rather than a clipping of the regular past participle form *pedido* (e.g. *'*pedo*', formed by removing the thematic vowel and *-d*). This leads Chagas de Souza (2007; 2011; 2015) to claim that they represent the emergence of a new syncretism, where or more two sets of values for morphosyntactic features are represented by a single inflected form, as defined in section 3.3.3 above. In this case, the syncretism would link the first person singular indicative form with the past participle, by analogy with inherited forms like *expulso* 'expelled /I expell' (irregular past participle, and first person singular present indicative). This is schematised in Table 16 below, summarising information from Chagas de Souza (2007).

	Infinitive	Translation	Regular PTCP	Irregular PTCP	1SG Pres. Indic.
Inherited double forms from Latin	ACEITAR	'accept'	*aceitado*	*aceito* (BP) *aceite* (EP)	*aceito*
	EXPULSAR	'expel'	*expulsado*	*expulso*	*expulso*
Extension within first conjugation: standard Portuguese	PAGAR	'pay'	*pagado*	*pago*	*pago*
Innovation within first conjugation: non-standard	CHEGAR	'arrive'	*chegado*	*chego*	*chego*
Innovation in other conjugations: non-standard	TRAZER	'bring'	*trazido*	*trago*	*trago*
	PEDIR	'ask for'	*pedido*	*peço*	*peço*

Table 16 Extension of athematic past participle forms

Schwenter et al. (2019) note that irregular past participles that are identical in form to the first person singular present indicative are more frequently used than those which are not, leading them to agree with Chagas de Souza that 'analogy' between the two forms is relevant.

The existence of these innovative past participles in Brazilian Portuguese shows divergence between Brazilian Portuguese and European Portuguese. Since my focus is on European Portuguese, I largely leave the Brazilian Portuguese innovative past participles to one side in this

chapter. However, interesting empirical evidence about the innovative double past participles exists, and I develop an analysis of these forms as part of my diachronic analysis in Chapter 5. Therefore I briefly set out here some further generalisations about the innovative double past participles made by linguists working in the Distributed Morphology (DM) framework (Lobato 1999; Scher, Lunguinho, and Rodero-Takahira 2013; 2014; Nevins 2015; Nevins and Rodrigues 2014).[54]

Nevins (2015) argues that there are differences in both 'morphophonological form and grammatical semantic interpretation' between regular and innovative irregular forms. He develops a DM analysis proposed by Lobato (1999) in which the past participle forms are differentiated by underlying formal features. In experiments conducted by Nevins using invented words, Brazilian Portuguese speakers preferentially used irregular participles in contexts with simpler argument structure (unaccusative verbs) compared with transitives, ditransitives and psych-verbs (in order of increasingly complex argument structure). Nevins proposes that the irregular form and the regular form are distinguished by the existence of an additional functional head in the regular form, projected by a theme vowel:[55] the irregular form is therefore associated with 'lighter' verbal argument structure (2015, 228). However, this DM-specific analysis has no explanation for why the vast majority of Portuguese verbs have only one past participle form (whether regular or irregular) which is used in all contexts, regardless of argument structure.

---

[54] See sections 2.1.2, 2.2.1 and 3.3.1 for brief comments on DM, including a reference to LFG-based work that implements DM.

[55] The status of theme vowels in DM is somewhat different from in other analyses. This distinction is discussed briefly in section 3.3.1.

Another perspective on the distribution of the Brazilian Portuguese innovative irregular past participles is given by Scher, Lunguinho and Rodero-Takahira (2013). They argue that these forms are used only in perfect contexts and are incompatible with passive contexts, unlike regular forms which are 'underspecified forms compatible with both perfect and passive sentences' (2013, 106). They give grammaticality judgements such as (121) and (122), showing a distinction in the grammaticality of innovative irregular forms between the active context with the auxiliary *ter*, and the passive context with the auxiliary *ser* (2013, 105):

(121) A Maria tinha comprado / compro frutas de manhã.
the Maria had boughtREGPTCP / IRREGPTCP fruits of morning
'Mary had bought fruits in the morning'

(122) As frutas foram compradas / *compras pela Maria.
the fruits were boughtREGPTCP / IRREGPTCP by.the Maria
'The fruits were bought by Maria'

However, the same authors later (2014, 47) present evidence that speakers also use them in passive contexts, as in (123):

(123) […] meu bilhete tinha sido compro
my ticket had been bought. IRREGPTCP
em uma tarifa normal.
in a tariff normal
'My ticket had been bought at a normal rate.'

This raises the possibility that for some speakers, the innovative irregular form has simply replaced the inherited regular form. Scher, Lunguinho and Rodero-Takahira (2014) admit that their data regarding the use of these innovative forms is unsatisfactory, as it is a collection of sporadic instances of these forms found across the internet, leaving the data open to claims that it contains performance errors or hypercorrections. Nonetheless, they argue that such forms need to be

accounted for, and present a DM analysis involving the presence or absence of particular formal features.

The empirical evidence about syntactic distinctions between the regular and irregular past participles in Brazilian Portuguese, including innovative irregular forms, is therefore contested. Further data about the innovative forms and their distribution are discussed in Chapter 5, where I give my own theoretical analysis of these forms.

4.3.4 Interim summary

As discussed in Chapter 2, I reject the DM conception of morphology in favour of a PFM analysis. In the rest of this chapter, I set out my LFG-PFM account of the Portuguese past participles. This analysis does not rely on functional heads or argument structure properties to distinguish between the regular and irregular past participle forms. Instead, for each Portuguese verb, the required past participle form or forms, associated with the correct morphological features, are produced in the PFM morphological component for insertion as whole words into the LFG-theoretic syntactic modules, according to the principle of Lexical Integrity.[56]

This LFG and PFM approach effectively captures the facts of the double past participle phenomenon, including the importance of lexical specification, and the potential for substantial variability across regional varieties, between individual speakers, and over time. More specifically, it allows an effective account of the relationship between the different functions and forms of the past participle for different Portuguese verbs, and makes explicit the relevant differences.

---

[56] As we have seen, past participles are often used in periphrases, which can be treated as multiword realisations of a single lexeme. I explain possible treatments of the past participle from this perspective in section 4.4.9.

Table 17 summarises the information that the formal LFG-PFM model needs to capture about the correspondence between past participle forms and functions in Portuguese, for three types of verbs: verbs with only a regular past participle, verbs with only an irregular past participle, and verbs with double past participle forms. Some examples of each type are given.

Type of verb	**Active type past participle**	**Passive type past participle**
REGULAR VERBS *cantar* 'sing', *comer* 'eat', *partir* 'break'	Regular *cantado, comido, partido*	Regular *cantado, comido, partido*
VERBS WITH ONLY AN IRREGULAR PAST PARTICIPLE *pagar* 'pay', *fazer* 'do', *abrir* 'open'	Irregular *pago, feito, aberto*	Irregular *pago, feito, aberto*
DOUBLE PAST PARTICIPLE VERBS *entregar* 'hand over', *acender* 'light, turn on', *exprimir* 'express'	Regular *entregado, acendido, exprimido*	Irregular *entregue, aceso, expresso*

Table 17 Forms and functions of the past participle for different types of verbs

With this summary of the data in place, the next section presents my formal theoretical account of the Portuguese double past participles.

4.4 A formal LFG-PFM account of double past participles

Having laid the foundations for a formal analysis of the different past participle forms in Portuguese, and in particular, their relationship with the verbal paradigm, I now present my analysis in the LFG-PFM theoretical framework. This section applies a number of extensions to standard PFM which help to capture the properties of double past participles in Portuguese.

I begin by discussing two key analyses of double past participles: as a morphological phenomenon, and as a syntactic phenomenon (section 4.4.1). Section 4.4.2 takes up the discussion of lexical category in relation to past participles, presenting two alternative theoretical approaches to the complexities discussed above in section 4.2.2. In section 4.4.3, I present my analysis of how past participles relate to the verbal paradigm.

The rest of section 4.4 develops the PFM rules required for this analysis. In section 4.4.4 I demonstrate how the analysis deals with lexical specification. Section 4.4.5 gives the PFM rules that produce the correct regular and irregular forms as required for Portuguese verbs. This presentation is extended in section 4.4.6, which includes worked examples to demonstrate how the rules apply, and in section 4.4.7, which shows how the analysis accounts for the morphological agreement of the passive past participles. Section 4.4.8 discusses how the rules capture the apparent syncretism between active and passive forms in most verbs. In section 4.4.9 I show how the forms produced in these rules for past participles can feed into existing models of periphrasis, accounting for the correct multiword paradigm cells for each verb. Finally, in section 4.4.10 I demonstrate how this morphological analysis of the past participle in periphrases fits alongside a syntactic/semantic LFG account of the changes in mapping between grammatical functions and argument positions for voice alternations, since the passive periphrasis is a key context for use of the past participle in Portuguese.

Throughout section 4.4, I consider the challenges of accounting for Portuguese double past participles from the formal theoretical perspective of the LFG-PFM framework proposed and

discussed in Chapter 2. So, for example, I assume that morphology and syntax are separate modules, with syntax operating over whole words (the Lexical Integrity Principle: see section 2.1.2), and that the lexical entry for each wordform includes fine-grained information about its morphological and syntactic features. The different elements of the LFG-PFM model thus combine to produce the whole wordforms and periphrastic forms, in each case paired with their sets of grammatical features, that can fulfil a role in the LFG model of the syntax.

4.4.1 Morphological and syntactic views of double past participles

Double past participles represent an interesting case study for the interaction of morphology and syntax in this theoretical framework. Regular Portuguese verbs, as well as most irregular Portuguese verbs, appear to have one past participle form which, in the LFG-PFM model, is produced in the morphological component, and occurs in sentences produced by the syntactic component. In double past participle verbs, two different past participle forms exist, whose distribution is divided between two sets of functions. It therefore must be established which part of the grammar accounts for the right form occurring in the right functions, for double past participle verbs.

One possible view is that the differences are purely morphological: a double past participle verb has two past participle forms, rather than one, meaning that it shows morphological overabundance, with a single slot in the paradigm being filled by two competing forms (Thornton 2011). Thornton (p361) gives a range of examples of overabundance from Italian, including:

(124) *vado* / *vo* 'go:1SG.PRS.IND', *faccio* / *fo* 'do:1SG.PRS.IND'

(125) -*bb*- / -*v*- forms of DOVERE 'must' (e.g. *devo* / *debbo* 'must:1SG.PRS.IND')

(126) Forms with/without diphthong ('*dittongo mobile*') in SEDERE (e.g. *siederò* / *sederò* 'sit:1SG.FUT')

There is a substantial degree of variation in the relative frequency of the two forms. For example, in Thornton's corpus, there are more than 10,000 token of *faccio*, and only 56 of *fo*; for the future tense of SEDERE, there are 854 tokens in *-ie-* and 129 tokens in *-e-*.

Canonical overabundance involves free variation between the two forms (Thornton 2011, 362), in which the two forms 'can be used interchangeably, with the choice of one or the other form subject to no condition'. Overabundance is found to a limited extent in the past participles of other Romance varieties, e.g. Standard Italian: see Thornton (2011) for a discussion of the forms *sepolto/seppellito* 'buried', although there are may be distinctions in register between the two forms. However, the Portuguese double past participles in the European examples discussed here are far from canonically overabundant: the distribution of the two forms is determined by their context, whether in the active periphrasis with *ter* 'have', or in passive periphrases, or other contexts.[57]

An alternative view of the double past participle verbs is that they reveal a broader fact about Portuguese, namely that rather than there being one past participle, there are in fact two separate participles, with different functions, which have different morphological forms in the double past participle verbs, while having the same form in other verbs. This can be seen as a 'morphomic' approach to the past participle: the fact that both functions take the same morphological form for the vast majority of verbs is an arbitrary morphological regularity (a morphome: see for example Aronoff 1994; Maiden 2016b). From this perspective, it is easy to incorporate the fact that passive context past participles agree in gender and number, whereas active context past participles do not

[57] An overabundance analysis may also be appropriate for some Portuguese varieties, where there is no active/passive distinction between the two forms, but a single speaker might use both forms. Varieties with no active/passive distinction are discussed in Chapter 5.

show agreement, as discussed in section 4.2.2.1. Similarly, the two past participle functions can be viewed as syncretised (see section 3.3.3 above).

In both the overabundance view and the syncretism/morphome view, the existence of two past participle forms is a morphological fact about the structure of the verbal paradigm. However, syntax is also relevant, because the context in which each past participle form is found (active vs. passive) is defined with reference to syntactic properties (or, to use Matthews's term, 'morphosyntactic features'), i.e. voice.[58] The alternation between active and passive voice is realised through differences in the alignment of semantic roles and grammatical functions, and associated with differences at multiple levels of syntactic structure: f-structure, c-structure, and in some conceptions of the LFG architecture, argument structure (see Dalrymple, Lowe, and Mycock 2019, chap. 9).

An effective account of the double past participles needs to include information about voice. This view can be reconciled with the 'syncretism' view of double past participles, recognising that the voice features of the active and passive periphrases can (for a subset of verbs) feed into the selection of different morphological forms: apart from the morphological coincidence that for many verbs, the same form is selected (albeit with differences in agreement), there is no reason to assume that the past participle is a single entity with only one form and one set of properties. The double past participle verbs provide evidence to the contrary for Portuguese.[59] The two past participle forms can be distinguished by the presence of the syntactic voice features ACTIVE and PASSIVE, which are relevant at f-structure.

---

[58] This may not be the case for Brazilian Portuguese: see section 4.3.3 above.

[59] Whether the same is true of other languages is a separate question which is beyond the scope of this thesis, but this work provides a framework for capturing such cases across languages.

This argument has previously been made by Maiden (2016b, 708), who argues that Portuguese double past participle verbs are 'probably the sole clear case in Romance where these different functions are distinguished'. As in many instances, this mismatch which is relatively marginal within the language reveals interesting linguistic properties that would otherwise go unnoticed, and which have great significance for our understanding of the different aspects of linguistic structure and their interactions.

Having briefly considered the relationship between syntax and morphology in the double past participle phenomenon, I next look at another of its challenging characteristics, namely the lexical category properties of past participles.

4.4.2 Theoretical approaches to the lexical category of past participles

As discussed in section 4.2, past participles do not fit neatly into one lexical category, as they may have both adjectival and verbal properties. They also have different properties when used with different functions in Portuguese, leading us to distinguish between *active type* past participles and *passive type* past participles. This section considers some theoretical approaches that can resolve the problems associated with the lexical category classification of past participles. I introduce the section by presenting Aarts's (2004; 2007) influential theory of 'gradient categories', and then discuss two theories that attempt to resolve the problems from an LFG and a PFM perspective respectively, before summing up my own approach.

Aarts (2004; 2007) argues for a 'gradient' analysis of category membership, in which certain sub-categories behave more like members of another category: some verbal forms may behave more like adjectives, and some adjectival forms may behave more like verbs, while still maintaining a separation between adjectives and verbs. Aarts argues for an analysis that 'allows for gradience both within and between categories, but does not do away with clearly delimited categorial boundaries' (Aarts 2004, 383).

At first glance, it might appear that a gradient analysis is incompatible with the formal frameworks of LFG and PFM, because these theories demand that words are members of a particular category. In PFM, realisation rules generally apply to words of one category.[60] As Lowe (2016, 405) notes, 'within LFG, there is no concept of intermediate categories: there is a finite inventory of distinct lexical categories, and every lexical word must belong to exactly one category'. However, from both perspectives, it is possible to implement Aarts' ideas, maintaining boundaries between categories by separating out differences in morphological and syntactic behaviour that are characteristic of different categories. I consider two approaches: Lowe (2016; 2019) in section 4.4.2.1 and Spencer (2013) in section 4.4.2.2.

4.4.2.1 An LFG approach: participles as a mismatched category

Various approaches to lexical categories are possible within LFG; see Nikolaeva and Spencer (2019, chap. 4) for an overview. Here I focus on the approach described by Lowe (2016; 2019) which looks mainly at phrase structure. This approach distinguishes between 'mixed categories' and 'mismatched categories'. Mismatches within internal syntax are indicative of mixed categories, while mismatches between morphosyntax, distribution and/or internal syntax are indicative of mismatched categories. The analysis rests on the insight that 'phrase structure rules distinguish different categories in terms of the types of specifiers, complements and adjuncts that they admit'. The analysis is illustrated (Lowe 2019) for the Arabic *masdar* (verbal noun) and the English 'Type C gerund' which has properties of a DP and of a VP.

Lowe argues that attributive participles are not a mixed category. They have mixed morphosyntactic behaviour, and their distribution is adjectival, but their internal syntax is purely verbal. He gives the following evidence (Lowe 2019, 31):

[60] An obvious exception is the Identity Function Default, which applies across all categories.

- 'Participles to transitive verbal stems may take objects, in the same case as, and entirely indistinguishable from, the object of a corresponding finite verb form', and:
- 'Participles may be modified by adverbs but never by adjectives.'

The mismatch between their morphological behaviour (adjectival) and their syntactic category (verbal) means that they are a mismatched category. The behaviour of 'mismatched categories' is analysed using Dalrymple's (2015) architecture for the morphology-syntax interface, which allows a distinction between morphosyntactic agreement properties and syntactic category. A form of one syntactic category can have morphosyntactic agreement properties typical of another category: morphological agreement properties do not influence the syntactic category of a word.[61]

This distinction is captured using the Description Function $\mathcal{D}$, which is defined in (127):

(127) $\mathcal{D}$ <Lexemic Index, m-features, category, f-description$_M$ >

The Description Function includes the function $\mathcal{D}_{cat}$ which specifies the c-structure category based on the LI and m-features. Typically the $\mathcal{D}_{cat}$ takes the following form:

(128) $\mathcal{D}_{cat}$ <LI, m-features, N> iff M-CAT:N ∈ m-features.

$\mathcal{D}_{cat}$ <LI, m-features, V> iff M-CAT:V ∈ m-features.

$\mathcal{D}_{cat}$ <LI, m-features, Adj> iff M-CAT:ADJ ∈ m-features. etc.

[61] Lowe gives the example of Russian past tense verbs, which diachronically derive from verbal adjectives, and synchronically have the agreement properties of adjectives (agreeing in gender and number instead of person and number) but are still unambiguously verbal by the diagnostic of internal syntax.

As with other mismatches, mismatched categories have a disruption to the normal mapping between morphological form and syntactic category. This is captured by specifying a different $\mathcal{D}_{cat}$ that applies to specific sets of LIs. In the following examples, LI_{adj} are lexical adjectives (those with 'lexical entries which are fundamentally adjectival, i.e. {HAPPY1, SAD1, TALL1, FIERCE1}') while LI_{vb} are 'the set of LIs associated with a verbal lexeme, i.e. {GO1, FIND1, LOVE1 …}' (Lowe 2016, 418).

(129) $\mathcal{D}_{cat}$ <LI_{adj}, m-features, Adj> iff M-CAT:ADJ ∈ m-features.

(130) $\mathcal{D}_{cat}$ <LI_{vb}, m-features, V[ptc]> iff M-CAT:ADJ ∈ m-features.

The crucial fact is that 'the specification in [example (130)] associates the m-feature M-CAT:ADJ with the syntactic category V, for a certain set of lexemes in the language' (p418). V[ptc] is a subcategory of V, which is proposed 'to effectively subdivide the category V in such a way that participle Vs can distribute differently from finite and other Vs' (p414).

In this analysis, subcategories like $V_{[ptc]}$ and $V_{[fin]}$ have different agreement properties, captured through a template in the lexical entry. For example, the subcategory $V_{[ptc]}$ has the agreement properties of an Adj, but is still classified syntactically as a (subcategory of) V. The template for 'morphological adjectivehood' (Lowe 2016, 417) applies to both lexical adjectives and participles, specifying what m-features morphological adjectives have:

(131) M-ADJ(_CASE, _NUM, _GEND)≡M-CAT:ADJ, M-CASE:_, M-NUM:_, M-GEND:_

This approach has been criticised in some detail by Nikolaeva and Spencer (2019, chap. 4) for several reasons, including that category membership as a verb has to be stipulated for participles, which takes away from the original intention to provide an analysis of participles as verbs (due to their internal syntax). In addition, Nikolaeva and Spencer remark that the suggested solution for Indo-European-type participles cannot easily be generalised to similar phenomena in other languages (other types of transpositions: see section 4.4.2.2). They suggest that attempts to resolve

this problem by stipulating the category membership of other similar forms, such as predicative adjectives, would lead to unwanted implications for c-structure, as a predicative adjective that heads a clause would remain a member of category A: extending this analysis means that we would have to permit phrase structure rules of the general form ZP → XP YP 'for more or less any X, Y, Z' (Nikolaeva and Spencer 2019, 168). Nikolaeva and Spencer argue that a version of Spencer's (2013) *Lexical Relatedness* is a more effective way of approaching mixed or mismatched categories. This is discussed in the next section.

4.4.2.2 A PFM approach: participles as transpositions

Spencer (2013) considers lexical categories in some detail, as part of his discussion of the full range of ways in which lexical items can be related, beyond simple inflection and canonical derivation, in his Generalised Paradigm Function Morphology.[62] I refer the reader to Chapter 3 above for an overview of the main concepts.

Transpositions, of which past participles are an example (p346), are forms which are 'transposed' from one lexeme to another category, such that they retain some features of the base lexeme, and have other properties associated with another category. Spencer also describes this as 'category shift without meaning change' (p136). As we have seen, past participles are closely related to the base lexeme category (verb) but have some properties of adjectives. In this view, transpositions are neither straightforwardly inflectional nor obviously derivational.[63]

---

[62] Spencer focuses on simplex forms. For a brief discussion of Spencer's treatment of periphrases from this perspective, see section 3.3.4.

[63] According to Beard (1995, 177), Szymanek (1988) put forward early work on transpositions. However, in his analysis, transpositions are a subset of derivational forms, and participles are treated as inflectional forms, rather than as transpositions. This points to the importance of being

As well as simple 'transpositions', Spencer distinguishes 'meaningful transposition' (transposition that includes semantic change, i.e. different FORM, SYN and SEM from the base), '*Angestellte(r)* nouns' (meaning-bearing transposition which preserves the inflectional properties of the base, i.e. altered SYN and SEM from the base), and 'm-inert transposition' (altered SYN). Spencer's category of 'meaningful transpositions' is useful for the analysis of Portuguese forms which are now purely adjectival and have developed semantically from the original past participle (transposition) that they derived from, such as *tinto* 'red' (see section 4.3.3). Although these might also be considered as derivations in synchronic terms, they are likely to have gone through a phase as meaningful transpositions, since at some point they would have retained a closer connection to the original lexeme, while having a difference in meaning.

Spencer notes that the literature on most of the possible types of lexical relatedness is inadequate, disguising the most interesting morphological facts and failing to account for the full range of possible properties. For LFG, in particular, Spencer states that (p300) 'There is no general account of transpositions in LFG. A number of proposals have been put forward for handling gerund-type constructions in English …. but there is no general account of participles, relational adjectives, and the like.'

In Spencer's analysis, a transposition differs in FORM and SYN from its base lexeme, but its SEM and LI are inherited from the base lexeme. This is schematised in (132), adapting the example for derivation given in section 3.3.2.[64]

---

clear about the meanings of these terms, as different meanings are intended by linguists in different theoretical contexts.

[64] Spencer (2013, 347) gives a more complex analysis for Latin past passive participles, which accounts for the voice and argument structure of the participle within the f_{syn} function.

(132) Where φ = {[VFORM:PPL]},

GPF(<WRITE, φ>) ≡

f_{form} (<WRITE, SN>) = *written*

f_{syn} (<WRITE, SN>) = Adjective

f_{sem} (<WRITE, SN>) = [PERSON(x), [WRITE(x,y)]]]

f_{li} (<WRITE, SN>) = WRITE

This analysis can be extended to a Portuguese past participle, such as *cantado* from *cantar* 'sing':

(133) Where φ = {[VFORM:PPL]},

GPF(<CANTAR, φ>) ≡

f_{form} (<CANTAR, SN>) = *cantado*

f_{syn} (<CANTAR, SN>) = Adjective

f_{sem} (<CANTAR, SN>) = [PERSON(x), [SING(x,y)]]]

f_{li} (<CANTAR, SN>) = CANTAR

A more fine-grained account of the FORM properties represented by the form 'cantado' and the syntactic properties I have represented here by 'Adjective' is of course necessary. Spencer (2013, 184–88) makes detailed proposals for the actual content of FORM, which includes information about stems and inflection class (i.e. morphological information). However, an in-depth exposition of this is beyond the scope of this chapter. The shorthand used here is a representation of the fact that a past participle may have the syntactic properties of an adjective, depending on the context of its use, while retaining a semantic connection to the base (verbal) lexeme.

4.4.2.3 Reconciling different approaches to lexical categories

So far in section 4.4.2 we have seen two contrasting approaches to dealing with the complex lexical category properties of past participles. Lowe's approach builds on LFG assumptions, but has a narrow focus on particular types of lexical relatedness. Spencer's approach, by contrast, takes PFM

as its theoretical starting point, and is designed to capture the behaviour of a large range of forms, for some of which there is no existing LFG analysis.

My analysis of the past participle forms does not depend on a particular theoretical analysis of lexical category. Up to a point, it is compatible with either Lowe's or Spencer's models, both of which clearly have useful insights. In the remaining part of this chapter, I follow Spencer in assuming that in general past participles are transpositions which share an LI with their base lexeme. This allows them to be included in the inflectional paradigm of the verb. Without taking on the entirety of Spencer's GPFM approach, I present a simplified system which produces transpositions (as part of the verbal paradigm) by the same blocks of PFM rules that derive purely inflectional forms. However, I consider active type past participles in Portuguese to be purely inflected forms. As morphosyntactic criteria are crucial in a PFM-based morphological system, and the active past participles do not have any distinctively adjectival morphology (FORM properties, to use Spencer's terminology), they do not qualify for treatment as transpositions. On the other hand, I treat both active periphrases and the passive periphrases as inflected forms.[65] This has implications for the paradigm models presented in section 4.4.3.

My approach as presented here focuses on the morphosyntactic aspects of the double past participle phenomenon. This leaves open the possibility of a further investigation of other facets of the lexical category of the past participle using Lowe's LFG approach, Spencer's GPFM approach, or alternative models.

[65] Passive periphrases could also be considered to be transpositions, with differences in SYN compared to the base lexeme due to the change in voice and mapping: see section 4.4.10.

Having shown how various aspects of the lexical category of past participles can be captured in a PFM-LFG analysis, I now move on to show how past participles fit into the wider verbal paradigm.

4.4.3 The place of past participles in the verbal paradigm

The concept of the paradigm is crucial to PFM, so this section is concerned with how past participles can be positioned in relation to the verbal paradigm. Past participles must be included in this paradigm because, as discussed above and following Spencer (2013), they share the LI with their base (verbal) lexeme. Figure 5 represents schematically the position of both active type and passive type past participles in the verbal paradigm in my analysis.

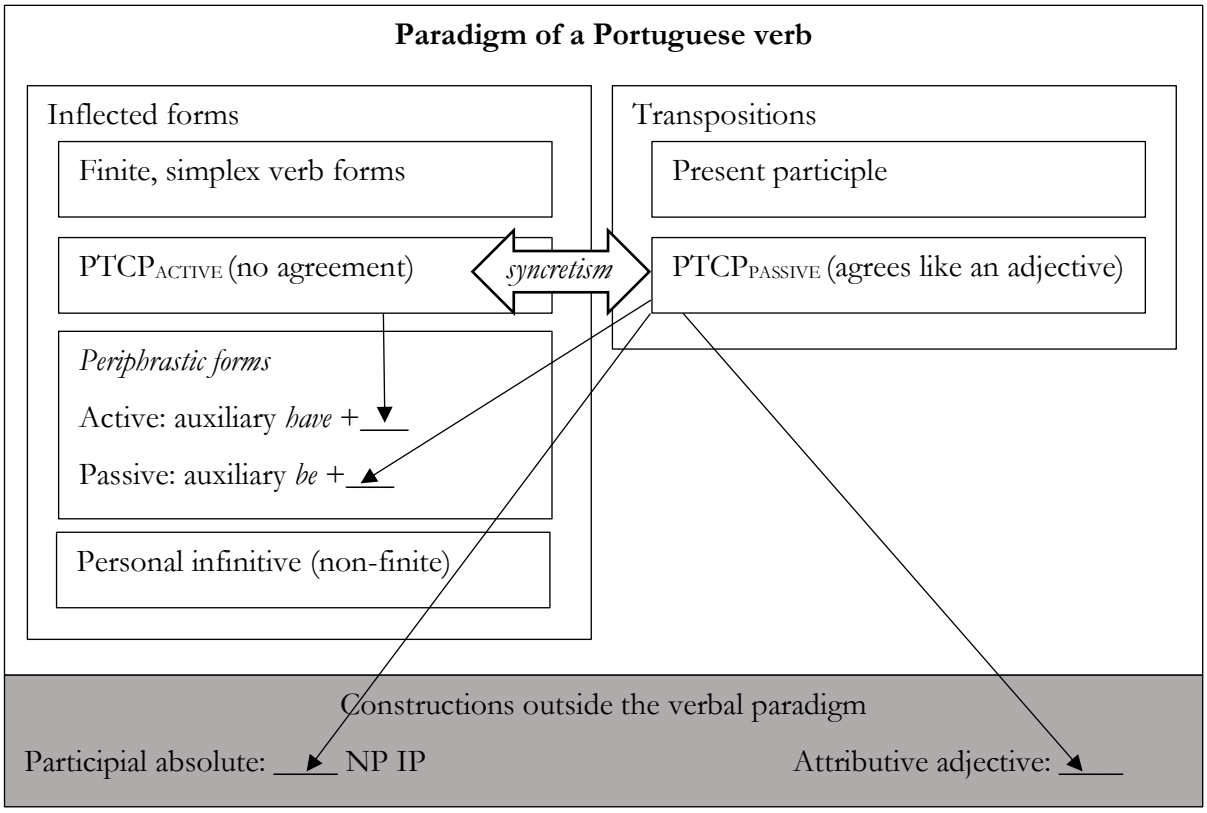


Figure 5 Past participles in the verbal paradigm

The figure represents my analysis that any lexical verb has two separate past participle forms, one of which has no agreement and occurs in active periphrases (PTCP$_{ACTIVE}$), and one of which agrees and occurs in passive periphrases (PTCP$_{PASSIVE}$). Both these past participle forms are in the verbal paradigm. However, the representation of their place in the paradigm reflects the differences in

morphological behaviour between the active type and the passive type past participles (see section 4.2). The active type past participle is an inflected form of the verb, while the passive type past participle is a transposition (see section 4.4.2). The arrows in this and similar figures should be taken to represent that forms can exist independently in the paradigm and occur in periphrases which are themselves part of the paradigm. The mechanism by which such forms come to appear correctly in the periphrases, as well as being generated by rule blocks, is discussed in section 4.4.9.

Figure 5 is not exhaustive, particularly in its representation of other forms. For example, 'finite, simplex verb forms' should be taken to represent the many paradigm cells filled by simple inflected forms. The representation of auxiliary verbs is also simplified: references to 'auxiliary' in the paradigms should be taken to include a referral to the separate paradigm of the auxiliary verb. A representation of the personal infinitive, which is non-finite but inflects for person and number, is included.[66]

Figure 6 annotates the same structure with forms of the regular verb *cantar* 'sing'. For *cantar*, as for the vast majority of Portuguese verbs, there is a syncretism linking the masculine singular of the passive type participle to the active type participle. This reflects the morphomic behaviour of the past participle, by which two functions that are in principle unrelated come to be realised by the same morphological form. In verbs that only have one form (regular or irregular), the syncretism applies.

[66] The personal infinitive is a form unique to Portuguese among the Romance languages (Ilari 2013, 59) in which the infinitive is inflected with the same endings as used for the future subjunctive. It is identical in form to the future subjunctive for many verbs, but in some verbs the stem used is different, as shown in Appendix I.

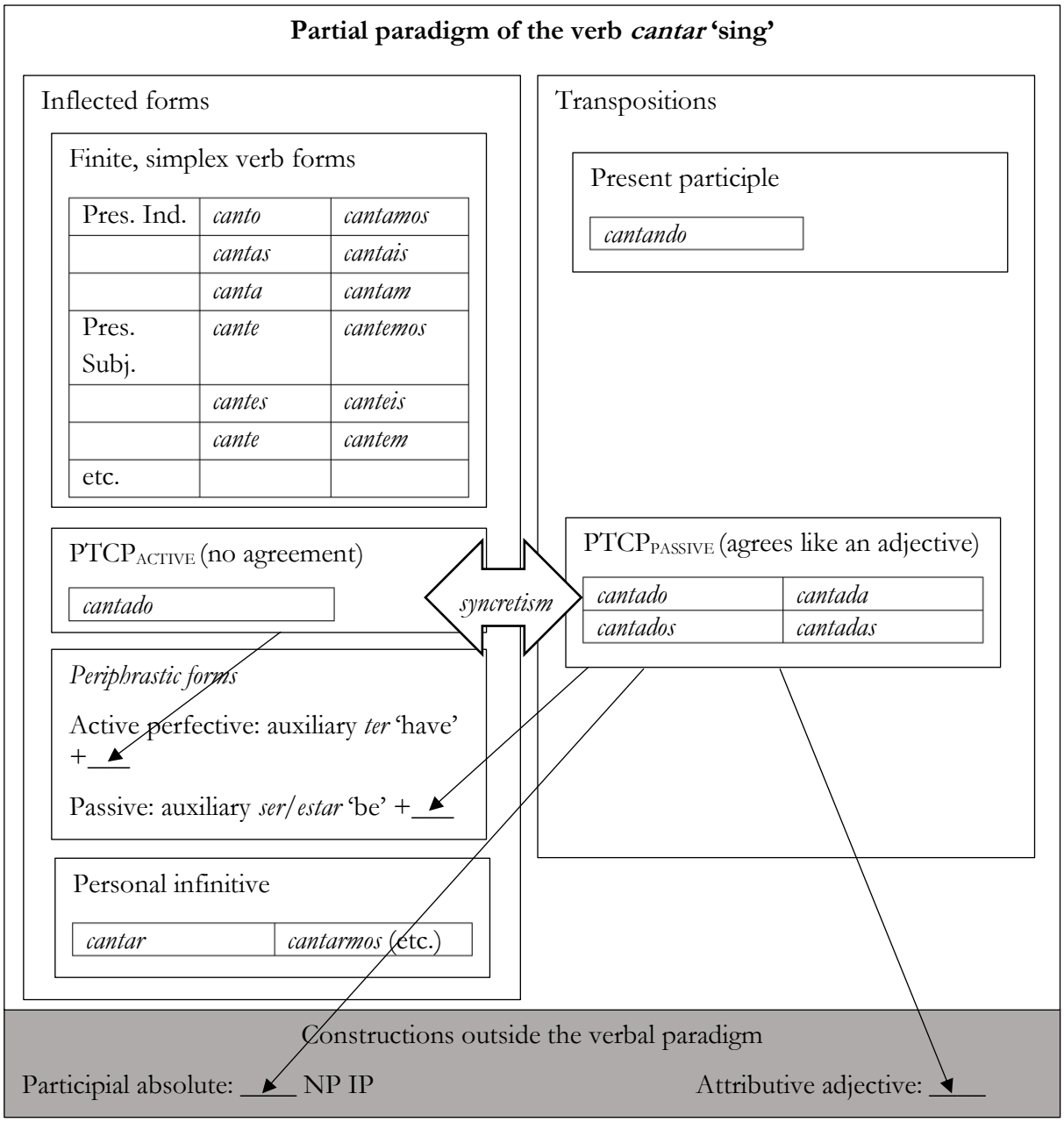


Figure 6 Partial paradigm of the regular verb *cantar* 'sing'

However, for 'double past participle' verbs, the normal syncretism does not exist: the masculine singular of the passive type participle is different in form from the active type participle.[67]

[67] A comparative note: this model is not appropriate for other Romance varieties, such as French or Italian, where the past participle can agree in active contexts. The status and functions of the past participle in these languages are clearly different from those of the past participle in

This is reflected in Figure 7, which shows the paradigm model annotated with forms of the double past participle verb *acender* 'light, turn on'. Section 4.4.8 discusses how the syncretism is captured in my PFM rules.

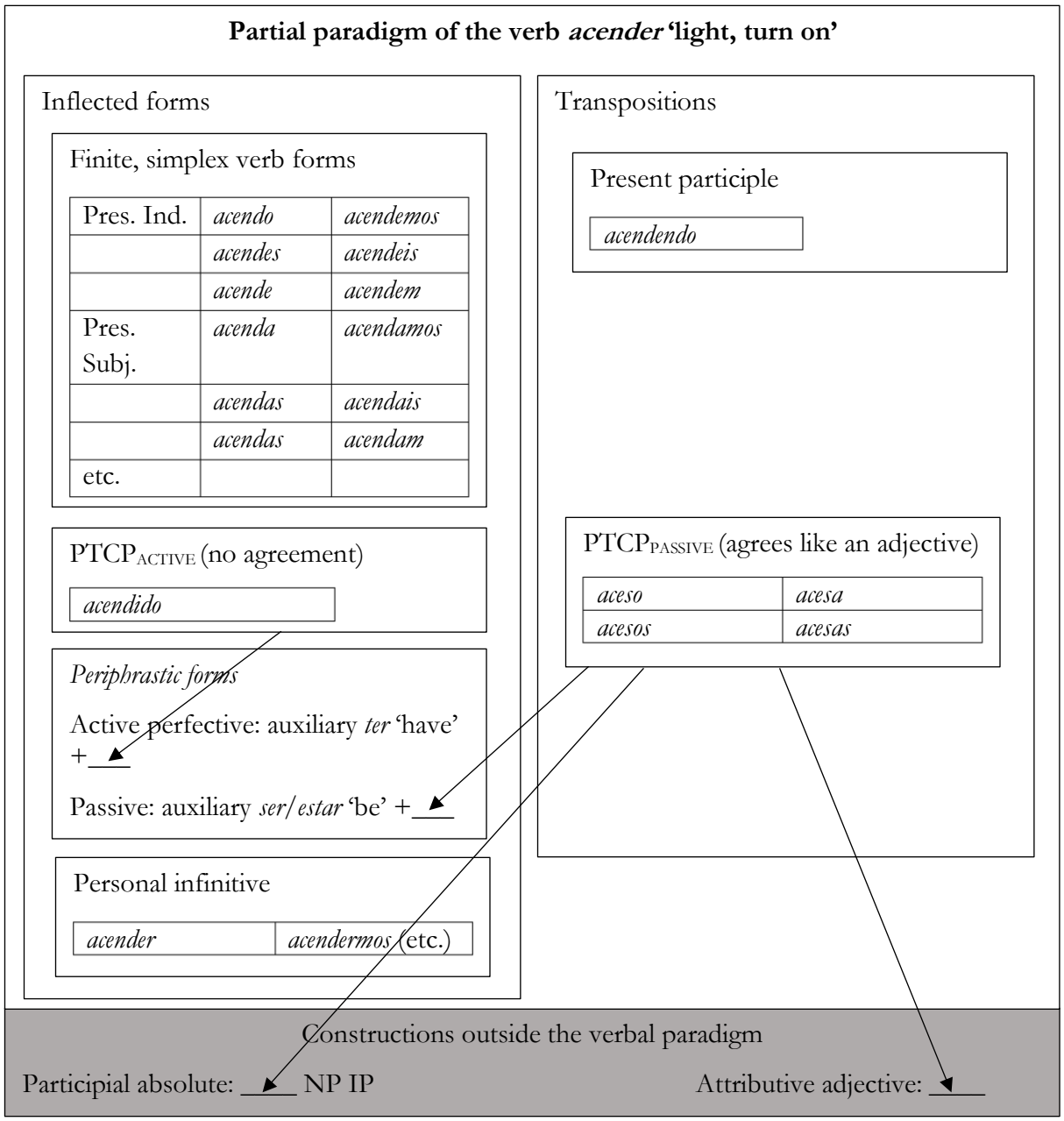


Figure 7 Partial paradigm of the double past participle verb *acender* 'light, turn on'

Portuguese, and the double past participle phenomenon does not exist in the same way in those varieties: where two forms exist, they are not divided up along the same lines.

Having seen the place of the different past participle forms in the verbal paradigm, I now move on to discuss how my analysis accounts for the lexical arbitrariness of the double past participle pattern, using inflection classes.

4.4.4 Lexical specification in the double past participle phenomenon

One of the challenges in accounting for double past participle verbs from a theoretical perspective is that the phenomenon is restricted to a fairly small subset of verbs, which varies depending on the language variety, both on a broad scale (Brazilian Portuguese versus European Portuguese) and also between speakers. Previous descriptive studies, as far back as Ali ([1908] 1957, first ed. 1908) and as recently as Schwenter et al. (2019), have demonstrated that the verbs that have two past participles are essentially arbitrary: there is no principled explanation for which verbs act like this, other than lexical specification. This means that an effective theoretical account must capture which verbs are included, without overgeneralising the pattern. Information about participation in the double past participle phenomenon must be captured at a lexical level.

From a diachronic perspective, the lexically arbitrary nature of the double past participle phenomenon is entirely expected. The trajectory of development of the different forms varies from verb to verb: a few verbs inherited two past participle forms from Latin, some inherited only an irregular form (and some of these later innovated a regular form), and the vast majority of regular verbs inherited a regular form (some of which later innovated an irregular form). Morphological irregularities of this kind are commonplace, and a high level of irregularity and unpredictability is expected at the morphological level, both in a broad cross-linguistic sense, and in the Romance context more particularly (see e.g. Maiden 2018). Chapter 5 considers the diachronic development of the past participles in more detail.

A level of arbitrariness or unpredictability in the morphological behaviour of the past participle is to be expected. Here I aim to capture this irregularity and arbitrariness in a theoretical framework,

allowing for lexical variation, while also accurately accounting for the systematic usage patterns seen for past participle forms of different verbs.

Another interesting question raised by the double past participle phenomenon is that for many of the verbs involved, their only irregularity is the fact that they have two past participle forms. In other respects they are entirely regular verbs of the relevant major inflection class (first, second or third). Similarly, some of the Portuguese verbs that have only an irregular past participle form are also otherwise entirely regular, such as *escrever* 'write'. It is necessary to strike a balance between recognising the lexical variation involved in the phenomenon (some verbs participate in it, and others do not), and the 'regularity' of the small class of verbs that participate.

Several inflection classes are required to capture the forms of Portuguese double past participles, as laid out in section 4.1.2 above. The rules which I propose to define these are 108-111. These rules use standard set theory notation to define the classes. **V** ∩ **Pt** is the class of verbs in Portuguese.

108. {ENTREGAR, GANHAR, GASTAR, LIMPAR, MATAR, PAGAR …} = **TruncPTCP** ⊂ **V** ∩ **Pt**

109. {ACENDER, EXPRIMIR, PRENDER, SUSPENDER …} = **S-IrregPTCP** ⊂ **V** ∩ **Pt**

110. {ABRIR, COBRIR, DIZER, ELEGER, ESCREVER, FAZER, MORRER, PÔR, VER …} = **T-IrregPTCP** ⊂ **V** ∩ **Pt**

111. **S-IrregPTCP** ∪ **T-IrregPTCP** ∪ **TruncPTCP** = **IrregPTCP** ⊂ **V** ∩ **Pt**

112. {ACENDER, PRENDER, EXPRIMIR …} = **DoublePTCP** ⊂ **V** ∩ **Pt**

The first three classes defined here, **TruncPTCP**, **S-IrregPTCP** and **T-IrregPTCP** respectively, include verbs with irregular past participles of a particular type: these are the truncated type

common to first conjugation verbs with irregular past participles, and the irregular past participles formed with a consonant /s/ or /t/ before the final vowel. The **TruncPTCP** class includes verbs like *entregar* 'hand over' (past participle *entregue*) and *limpar* 'clean' (past participle *limpo*).[68] An example of a verb in the class **S-IrregPTCP** is *acender* 'light, turn on', which has the past participle *aceso*; an example for **T-IrregPTCP** is *abrir* 'open', with past participle *aberto*.

These three classes combined include all verbs with an irregular past participle. We therefore define the class **IrregPTCP** as the union of these three subclasses (111).[69]

The **DoublePTCP** class includes all verbs that demonstrate the double past participle phenomenon, i.e. those that have both regular and irregular past participle forms (112).

The classes and rules are set up to derive the correct forms for each verb depending on its inflection class membership. Verbs may enter or leave these classes over time, as speakers reanalyse the verb, innovating new forms or failing to acquire existing forms. For example, some verbs such

---

[68] An adaptation of this analysis to account for Brazilian Portuguese innovative past participle forms, which have morphological properties in common with this class of past participles, is given in Chapter 5.

[69] Many, but not all, of the verbs in the **IrregPTCP** class specify a basic stem root l which is used in the formation of the irregular past participle form. This is similar to the strategy used in Taylor's (2012) extension of PFM, where inflection classes are based on the specification of particular stems by classes of lexemes. For example, her analysis includes a class of verbs called **PYTA** (*pretérito y tiempos afines*, 'perfect and related tenses') which specify the **PYTA** stem, stem root k, which is selected in the formation of certain tenses. The role of stem roots in my analysis is discussed in the following sections.

as *gastar* 'spend' and *pagar* 'pay' used to have a regular past participle which has now been lost, thereby stopping being members of the **DoublePTCP** class, as Chapter 5 discusses.

The PFM analysis given here represents various pieces of information in tables of lexical data (see Appendix III). The theoretical status of this lexical data in the LFG-PFM model may appear opaque. However, it can be straightforwardly integrated with Dalrymple's (2015) model of the syntax-morphology interface. In that model, lexemic entries include a slot for 'root and idiosyncratic stem forms', which is clearly the correct place for basic stem roots to be listed. Complex stem roots do not need to be lexically specified because they are defined through stem formation rules.

Information about theme vowels, presented here as a theme vowel array, may also be represented in the lexemic entries, since this is information that is relevant to the lexeme as a whole. At first glance, this dramatically increases the amount of information that needs to be specified in each lexemic entry. However, the theme vowel array of most verbs is fully predictable from its inflection class (whether first, second or third). The exceptions to this can be specified in the lexicon: for example, TV[7] is normally identical to TV[1], but could be listed in the lexemic entry as ∅ for the three verbs *dizer* 'say', *fazer* 'do' and *trazer* 'bring', while TV[8] is generally /o/ (=TV[10]), but can be listed in the lexemic entry for *aceitar* 'accept' and *entregar* 'hand over' as /e/. This represents a kind of reindexing (see section 3.3.1) which reduces redundancy across the lexicon as a whole while enabling lexeme-specific exceptions to be captured appropriately.

A more complete discussion of lexemic entries in the context of the double past participles is in 5.3.5, where I also discuss how diachronic changes in lexeme-level information can be represented at this level of analysis.

In the next section I present the basic PFM rules that produce the correct morphological forms of the past participle for Portuguese verbs, and illustrate their application in worked examples in section 4.4.6.

4.4.5 PFM rules for past participles as complex stems

This section describes the PFM rules that give the correct forms of the past participle, which are presented in Appendix III as part of the full verbal morphology rules for Portuguese. These rules take their lead from Taylor's (2012) analysis of Portuguese verb morphology, which uses the extensions to the standard PFM1 rules put forward in her thesis, namely regarding stem formation and selection, and theme vowel arrays. Taylor does not account for the double past participle phenomenon, generating only one past participle form per verb. The rules I propose therefore include specific new rules which build on Taylor's account in order to provide analyses of all kinds of irregular past participles, and the double past participle phenomenon. In Appendix III, the rules I have introduced for past participles are marked in green.[70]

My PFM rules provide for each verb both a past participle used in the active (which has the vform feature: ppl and the voice feature: active) and a past participle used in the passive (vform: ppl and voice: passive). These forms are distinguished by their specification for the voice feature. While each verb has a slot in the paradigm for each of these two functions, there are differences in the actual morphological forms that fill these paradigm slots. These different forms, and their distribution within the paradigm, are the key elements of the double past participle phenomenon.

Taylor (2012) analyses past participles as 'complex stems', that is, stems which are formed from basic stems. This concept is discussed in more detail in section 3.3.1 above. The general structure of a complex stem is:

(134) $(\text{Stem root} + \text{theme vowel})_{\text{basic stem}} + \text{extra phonetic material} + \text{theme vowel}$

[70] Some minor corrections to Taylor's rules have also been made. I thank her for her help in discussing these corrections and developments.

My analysis also uses this conception of the past participle, forming both the past participle used in the active and the past participle used in the passive as complex stems.

The PFM rules capture at the level of 'stem formation rules' the differences in possible morphological patterns for the past participles discussed in section 4.1.2 and demonstrated in Table 10. The groups of verbs that can be identified are summarised in Table 18.

Verb	Regular PTCP	Irregular PTCP	Pattern
Regular verbs			**stem – TV[6] – d – o**
CANTAR	*cantado*	-	X a d o
COMER	*comido*	-	X i d o
PARTIR	*partido*	-	X i d o
Verbs with truncated irregular past participle			**stem – e OR o**
ACEITAR	*(aceitado)*	*aceite*	X e
ENTREGAR	*entregado*	*entregue*	X e
GANHAR	*(ganhado)*	*ganho*	X o
GASTAR	*(gastado)*	*gasto*	X o
LIMPAR	*(limpado)*	*limpo*	X o
PAGAR	*(pagado)*	*pago*	X o
Verbs with irregular past participle in -t-			**stem – t – o**
ABRIR	-	*aberto*	X t o
COBRIR	-	*coberto*	X t o
DIZER	-	*dito*	X t o
ELEGER	*elegido*	*eleito*	X t o
ESCREVER	-	*escrito*	X t o
FAZER	-	*feito*	X t o
MORRER	*morrido*	*morto*	X t o
PÔR	-	*posto*	X t o
VER	-	*visto*	X t o
Verbs with irregular past participle in -s-			**stem – s – o**
EXPRIMIR	*exprimido*	*expresso*	X s o
ACENDER	*acendido*	*aceso*	X s o
PRENDER	*prendido*	*preso*	X s o
SUSPENDER	*suspendido*	*suspenso*	X s o

Table 18 Summary of morphological patterns of the past participle

The stems referred to in the column 'Pattern' are basic stem roots (represented here by X). As will become clear, these are sometimes the regular stem root g, and sometimes the irregular stem root l. The regular stem root g is the normal stem used in the inflection of most forms of the verb, such as /kant/ for the verb *cantar* 'sing'. The irregular stem root l is defined by verbs whose irregular

past participle is built on a different form, such as the form /aber/ for the verb *abrir* 'open' (irregular past participle *aberto*).

Forms given in brackets in Table 18 are attested but not generally preferred by native speakers of the modern language in any context. For a full view of the past participle, however, it is valuable to be able to easily derive these forms, which is in fact trivial, given their regularity when viewed alongside regular verbs. As we have seen, there is substantial variation in the membership of the class of double past participle verbs. For example, some European Portuguese speakers may still use the regular form *aceitado* from *aceitar* 'accept' in active contexts, rather than the irregular form *aceite*. The rules given here enable the correct forms to be defined according to whether these verbs are members of the inflection class **DoublePTCP** at a particular stage of the language, for a particular language variety, or for a particular speaker, and so on.

4.4.5.1 General scheme of rule blocks

The rules involved in the formation of past participles are set out in the following series of sections, along with an explanation of how they apply to produce the different patterns seen in terms of form and function. The PFM rules in full and the lexical data (arrays that show each verb's basic stem roots and theme vowels)[71] are given in Appendix III.

First, however, I present a general overview of the order and relationships of the rule blocks involved. The following order represents how the rules are presented in Appendix III, with a brief description of the function of each type of rule.

Verb classes

Defines membership of verb classes

[71] This can be thought of as PFM notational shorthand for information stored in each verb's lexical entry, as described above in this section.

Stem formation rules

Block 0 (BSR_0) 'Stem base selection rule': defines stem roots, and specifies which basic stem is used in the formation of the complex stem

Block 1 (SR_1) 'Stem formation rule': concatenates extra phonetic material as required for the formation of a complex stem

Stem selection rules (SSR_0)

Specify the stem root and theme vowel used in the formation of the stem (whether basic or complex) used in a particular context (e.g. 'infinitive')

These include basic stem selection rules and complex stem selection rules, which are in competition, with only one rule being relevant to the inflection of a given form

Whole word syncretism rules (RR_0)

Identify regular patterns of syncretism between whole wordforms

Realisation rules

Block 1 (RR_1): accounts for tense/mood endings and person/number agreement of some forms

Block 2 (RR_2): accounts for person/number agreement

When applying the rules to give fully inflected forms, we start by considering which stem is selected (looking at the stem selection rules), before working out how the stem is formed (using the stem formation rules), and finally progressing to the realisation rules. This process is described in more detail in the rest of section 4.4.5 and exemplified through worked examples in section 4.4.6.

4.4.5.2 Stem formation rules: Block 0

As stated above, the past participle is formed as a complex stem. Taylor's analysis of Portuguese gives only one past participle form for each verb, which is based on stem root c. In that analysis, complex stem roots are identified by one of the letters a-f, while basic stem roots are identified by the letters g-l. My analysis of Portuguese retains c as the complex stem root for regular past

participle forms (whether active or passive in function) and introduces the symbol ɐ for the stem root of irregular past participle forms.[72]

Rules 122 and 123 describe the formation of the stem roots c and ɐ. According to these rules, c is composed of the basic stem root g plus theme vowel [6], while ɐ is stem root l (which is only specified for verbs with an irregular past participle form) plus theme vowel [9]. Theme vowel [9] is defined as a zero for all verbs. The purpose of the zero theme vowel is to disambiguate a stem root from a stem, because in Taylor's model all stems end in a theme vowel (section 3.3.1); this is not intended to imply that there is a 'phonetically empty' vowel in this position.

122. $BSR_0,\{c\},\mathbf{V} \cap \mathbf{Pt}(<\chi, \iota>) =_{def} <[stemg[6]], \iota>$

123. $BSR_0,\{ɐ\},\mathbf{V} \cap \mathbf{Pt}(<\chi, \iota>) =_{def} <[steml[9]], \iota>$

4.4.5.3 Stem formation rules: Block 1

The stem formation rules in block 1 are responsible for the portion of the complex stem labelled as 'extra phonetic material' above. This phonetic material is the consonant /d/, /t/ or /s/ that precedes the final theme vowel in the past participle, according to whether the stem being formed is for the regular past participle (/d/), or the irregular past participle of a verb in the class **T-IrregPTCP** (/t/) or **S-IrregPTCP** (/s/). The irregular past participles of verbs in the class **TruncPTCP** do not have any extra phonetic material here, which is accounted for by rule 129. There is little competition in these rules of the rule block, as they all apply to different subsets of verbs.

---

[72] This notational decision was made to reduce the amount of re-lettering of stem roots: there are no unused letters in the alphabet between a and f, and it was unappealing to rename all the roots g-l as well.

126. SR_1,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/d/], ι>for all i

127. SR_1,{ɐ},**T-IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/t/], ι> for all i

128. SR_1,{ɐ},**S-IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/s/], ι> for all i

129. SR_1,{ɐ},**TruncPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi], ι>for all i

4.4.5.4 Complex stem selection rules

Rules 143-145 determine the selection of the complex stems c and ɐ. By the normal application of PFM rules, rule 143 is the most general rule, which states that morphological forms with the property {vform: ppl} are defined as stem c combined with theme vowel [10]. Rules 144 and 145 are *narrower* than 143 and therefore apply to the relevant types of verbs: rule 144 applies to verbs in the **IrregPTCP** class, and rule 145 to the formation of past participles with the feature 'voice: active' for verbs in the **DoublePTCP** class.

143. SSR_0,{vform: ppl},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemc [10]], σ>

144. SSR_0,{vform: ppl}, (**IrregPTCP**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemɐ [8]], σ>

145. SSR_0,{vform: ppl, voice: active}, (**DoublePTCP**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemc [10]], σ>

These rules are also responsible for the addition of the appropriate final theme vowel to the past participles. This is generally theme vowel [10], which is defined as /o/ for all verbs. However, some irregular past participle forms (namely *aceite* from *aceitar* 'accept' and *entregue* from *entregar* 'hand over') end in /e/ rather than /o/. For this reason, I also define theme vowel [8] for all verbs; it is /o/ for all verbs other than those with an irregular past participle ending in /e/.

4.4.5.5 Realisation rules

In Taylor's conception of PFM, realisation rules apply to stems formed and selected in earlier rule blocks. Few realisation rules are relevant to past participles; the relevant ones are discussed in more depth in section 4.4.7, which deals with the number and gender of passive past participles.

Importantly, no realisation rules exist which apply to forms with the vform: ppl feature and specify a person, tense, aspect or mood feature, capturing the non-finiteness of these forms. This is similar to the approach taken for the infinitive, which generally does not agree, but can agree in person and number when inflected as the personal infinitive in Portuguese. This approach means that it is not necessary to include a feature for finiteness in verbal rule blocks for Portuguese, which could be seen as an advantage because it removes an unnecessary feature in the morphological system. On the other hand, the feature of finiteness may be desirable from a syntactic perspective: existing LFG analyses of various languages include finiteness as a binary feature (Butt et al. 1999) or as a value of the feature VFORM (Börjars, Nordlinger, and Sadler 2019, 100) or VTYPE (Dalrymple, Lowe, and Mycock 2019, 63). However, there is no need for the syntactic feature to have a morphological correlate relevant to the PFM analysis.

4.4.6 Worked examples

Having explained above the rules I propose for the past participles, I next give some worked examples. In each case, the input is a verb (e.g. *acender* 'light, turn on') and the specification 'vform: ppl' and either 'voice: active' or 'voice: passive'.

For ease of reading, each example is set out on a separate page.

4.4.6.1 Past participle form of regular verb cantar *(active or passive)*

The verb *cantar* 'sing' is in the inflection class **C1**, and has only a regular past participle, so it is not in the class **IrregPTCP** or **DoublePTCP**. Thus the relevant stem selection rule is:

143. SSR_0,{vform: ppl},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemc [10]], σ>

This tells us that the form is built on the basis of stem c, and has theme vowel [10].

To find out what stem c is, we look at the stem formation rules, starting with Block 0. The applicable rule from Block 0 is:

122. BSR_0,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[stemg[6]], ι>

This tells us that the form is built on stem root g, and has theme vowel [6].

We look up stem root g and theme vowel [6] in the lexical data. These are given as /kant/ and /a/ respectively, giving us the output /kanta/.

We now proceed to Block 1. The applicable rule is:

126. SR_1,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/d/], ι>for all i

This tells us to concatenate the phonetic material /d/, giving us the output /kantad/.

Having formed stem c, we complete the formation of the past participle by returning to the stem selection rule, which selects theme vowel [10]. This is defined for all verbs in the lexical data as /o/, giving us the final form /kantado/, as desired (orthographic *cantado*).

4.4.6.2 Past participle form of regular verb comer *(active or passive)*

The verb *comer* 'eat' is in the inflection class **C2**, and is not in the class **IrregPTCP** or **DoublePTCP**. Thus the relevant stem selection rule is:

143. SSR_0,{vform: ppl},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemc [10]], σ>

This tells us that the form is built on the basis of stem c, and has theme vowel [10].

To find out what stem c is, we look at the stem formation rules, starting with Block 0. The applicable rule from Block 0 is:

122. BSR_0,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[stemg[6]], ι>

This tells us that the form is built on stem root g, and has theme vowel [6].

We look up stem root g and theme vowel [6] in the lexical data. These are given as /kom/ and /i/ respectively, giving us the output /komi/.

We now proceed to Block 1. The applicable rule is:

126. SR_1,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/d/], ι>for all i

This tells us to concatenate the phonetic material /d/, giving us the output /komid/.

Having formed stem c, we return to the stem selection rule, which selects theme vowel [10]. This is defined for all verbs in the lexical data as /o/, giving us the final form /komido/, as desired (orthographic *comido*).

4.4.6.3 Active past participle form of double past participle verb acender

The verb *acender* 'light, turn on' is in the classes **C2**, **IrregPTCP, DoublePTCP** and **S-IrregPTCP**. Since we are forming the active past participle, the relevant stem selection rule is:

145. SSR_0,{vform: ppl, voice: active}, (**DoublePTCP**) ∩ **Pt**(<χ, σ>) $=_{def}$ <[stemc [10]], σ>

This tells us that the form is built on the basis of stem c, and has theme vowel [10].

To find out what stem c is, we look at the stem formation rules, starting with Block 0. The applicable rule from Block 0 is:

122. BSR_0,{c},**V** ∩ **Pt**(<χ, ι>) $=_{def}$ <[stemg[6]], ι>

This tells us that the form is built on stem root g, and has theme vowel [6].

We look up stem root g and theme vowel [6] in the lexical data. These are given as /asend/ and /i/ respectively, giving us the output /asendi/.

We now proceed to Block 1. The applicable rule is:

126. SR_1,{c},**V** ∩ **Pt**(<χ, ι>) $=_{def}$ <[χi/d/], ι>for all i

This tells us to concatenate the phonetic material /d/, giving us the output /asendid/.

Having formed stem c, we return to the stem selection rule, which selects theme vowel [10]. This is defined for all verbs in the lexical data as /o/, giving us the final form /asendido/, as desired (orthographic *acendido*).

4.4.6.4 Passive past participle form of double past participle verb acender

The verb *acender* 'light, turn on' is in the classes **C2**, **IrregPTCP, DoublePTCP** and **S-IrregPTCP**. Since we are forming the passive past participle, rule 145 (for active past participles) does not apply. Rule 144 is narrower than rule 143, so the relevant stem selection rule is:

144. SSR_0,{vform: ppl}, (**IrregPTCP**) ∩ **Pt**(<χ, σ>) $=_{def}$ <[stemɐ [8]], σ>

This tells us that the form is built on the basis of stem ɐ, and has theme vowel [8].

To find out what stem ɐ is, we look at the stem formation rules, starting with Block 0. The applicable rule from Block 0 is:

123. BSR_0,{ɐ},**V** ∩ **Pt**(<χ, ι>) $=_{def}$ <[steml[9]], ι>

This tells us that the form is built on stem root l, and has theme vowel [9].

We look up stem root l and theme vowel [9] in the lexical data. These are given as /ase/ and ∅ respectively, giving us the output /ase/.

We now proceed to Block 1. The applicable rule is:

128. SR_1,{ɐ},**S-IrregPTCP** ∩ **Pt**(<χ, ι>) $=_{def}$ <[χi/s/], ι> for all i

This tells us to concatenate the phonetic material /s/, giving us the output /ases/.

Having formed stem ɐ, we return to the stem selection rule, which selects theme vowel [8]. This is defined in the lexical data as /o/, giving us the final form /aseso/, as desired (orthographic *aceso*).

4.4.6.5 Passive past participle form of double past participle verb entregar

The verb *entregar* 'hand over' is in the classes **C1**, **TruncPTCP** and **DoublePTCP**. The relevant stem selection rule is:

144. SSR_0,{vform: ppl}, (**IrregPTCP**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemɛ [8]], σ>

This tells us that the form is built on the basis of stem ɛ, and has theme vowel [8].

To find out what stem ɛ is, we look at the stem formation rules, starting with Block 0. The applicable rule from Block 0 is:

123. BSR_0,{ɛ},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[steml[9]], ι>

This tells us that the form is built on stem root l, and has theme vowel [9].

We look up stem root l and theme vowel [9] in the lexical data. These are given as /entreg/ and ∅ respectively, giving us the output /entreg/.

We now proceed to Block 1. The applicable rule is:

129. SR_1,{ɛ},**TruncPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi], ι>for all i

This tells us that the form is not altered, giving us the output /entreg/.

Having formed stem ɛ, we return to the stem selection rule, which selects theme vowel [8]. This is defined in the lexical data as /e/, giving us the final form /entrege/, as desired (orthographic *entregue*).

4.4.6.6 Past participle form of verb with only an irregular past participle, abrir *(active or passive)*

The verb *abrir* 'open' is in the classes **C3**, **IrregPTCP** and **T-IrregPTCP**. The relevant stem selection rule is:

144. SSR_0,{vform: ppl}, (**IrregPTCP**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemɛ [8]], σ>

This tells us that the form is built on the basis of stem ɛ, and has theme vowel [8].

To find out what stem ɛ is, we look at the stem formation rules, starting with Block 0. The applicable rule from Block 0 is:

123. BSR_0,{ɛ},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[steml[9]], ι>

This tells us that the form is built on stem root l, and has theme vowel [9].

We look up stem root l and theme vowel [9] in the lexical data. These are given as /aber/ and ∅ respectively, giving us the output /aber/.

We now proceed to Block 1. The applicable rule is:

127. SR_1,{ɛ},**T-IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/t/], ι> for all i

This tells us to concatenate the phonetic material /t/, giving us the output /abert/.

Having formed stem ɛ, we return to the stem selection rule, which selects theme vowel [8]. This is defined in the lexical data as /o/, giving us the final form /aberto/, as desired (orthographic *aberto*).

4.4.6.7 Interim summary

So far I have presented the PFM rules required to produce the correct form for the active past participle, and also produce the correct masculine and singular form for the passive past participle. As I discuss in the next section, though, we also need rules to account for the number and gender agreement of the passive past participle.

4.4.7 Number and gender agreement of the passive past participle

The passive past participle shows morphological agreement in number and gender, which can be conceived of as a 'paradigm with a paradigm' (Spencer 2017). The regular set of inflectional endings that makes up this paradigm was shown in Table 13, repeated here as Table 19.

	Masculine	**Feminine**
Singular	*-o*	*-a*
Plural	*-os*	*-as*

Table 19 Number and gender agreement in Portuguese

Table 20 gives examples of passive past participles with number and gender agreement. The past participle of the regular verb *levar* 'carry' is shown first, then the irregular past participle of the verb *acender* 'light, turn on'. Finally, the verb *entregar* 'hand over' is shown, which follows a slightly different pattern (which is shared by the verb *aceitar* 'accept'). The final vowel in the singular form is /e/ in both the masculine and the feminine.

	Masculine	**Feminine**	**Masculine**	**Feminine**	**Masculine**	**Feminine**
Singular	*levado*	*levada*	*aceso*	*acesa*	*entregue*	*entregue*
Plural	*levados*	*levadas*	*acesos*	*acesas*	*entregues*	*entregues*

Table 20 Examples of passive past participles with number and gender agreement

The rules to provide number and gender agreement are simple compared with the large number of rules required to account for the many features which are marked on finite verbs such as person, number, tense and mood. In Taylor's (2012) model, the past participle is a complex stem, which means it can act as input into realisation rules. I propose the following realisation rules to account for gender and number agreement respectively.

170. RR_1,{vform: ppl, agr: {gend: fem},**V** ∩ **Pt**(<χ/o/, σ>) $=_{def}$ <[χ_i/a/], σ> for all i

179. RR_2,{vform: ppl, agr: {num: pl},**V** ∩ **Pt**(<χ, σ>) $=_{def}$ <[χ_i/s/], σ> for all i

Both rules apply to forms with the vform feature ppl (past participle), and are part of the realisation rules for verbs (**V** ∩ **Pt**). The rules will only apply when an 'agr' feature (either gender or number) is required, which prevents them applying erroneously to active past participles. The first rule appears in Block 1 of realisation rules, and the second in Block 2, reflecting the fact that gender agreement occurs closer to the stem than number agreement for these forms, so it is appropriate for gender agreement markers to be added in Block 1 followed by number in Block 2. The first rule specifies that for past participles which are feminine, the final /o/ in the stem is replaced by /a/. This rule is formulated in a non-standard way; this is an attempt at dealing with non-concatenative morphology in essentially the same way as concatenative morphology. Normally in PFM an alteration of this kind should be dealt with in the stem selection rules, since the theme vowel /o/ is part of the (complex) stem. However, for agreeing (passive) past participles, the /o/ and /a/ clearly mark masculine and feminine gender, rather than appearing as a default form.

A possible solution to this problem would be to suggest that, as the passive past participle is a transposition, one of the changes to its morphological properties in the 'process' of being transposed from verb to adjective is that the final vowel of the complex stem is reanalysed as a marker of gender. In section 4.4.2.3 I proposed that the passive past participle is in some respects better treated as an adjective, rather than as a verb, because of its transpositional qualities. In this case, the stem of the adjective should not include the final vowel, in line with other adjectives; this could be seen as a reanalysis of the word structure associated with transposition. Then the relevant realisation rules could be stated more generally as applying to adjectives, rather than verbs:

170'. RR_1,{agr: {gend: masc},**Adj** ∩ **Pt**(<χ, σ>) $=_{def}$ <[χ_i/o/], σ> for all i

170". RR_1,{agr: {gend: fem},**Adj** ∩ **Pt**(<χ, σ>) $=_{def}$ <[χ_i/a/], σ> for all i

179'. RR_2,{agr: {num: pl},**Adj** ∩ **Pt**(<χ, σ>) $=_{def}$ <[χ_i/s/], σ> for all i

This alternative has some attractive qualities. However, it also opens up some additional questions. For example, what is the status of the vowel /e/ at the end of the forms *entregue* and *aceite*? In addition, the process of transposition from verb to adjective is not encoded in the PFM rules I have given in the same way it would be within GPFM. Ideally, the passive past participle form should be verbal for the purposes of stem selection (with the feature vform: ppl) but adjectival for the realisation rules. Further to this, an optimal model of Portuguese would use the same realisation rules (with extensions) for both adjectival and nominal forms.

Providing thorough rules for adjectival and nominal morphology in Portuguese is well beyond the scope of the thesis, and I do not develop this further in this work. For reasons of simplicity, I continue deriving the passive past participle as a verbal form in the realisation rules, while accounting for its gender and number agreement properties through the rules proposed above.

The application of these rules is so simple as not to require much exemplification, but a brief example is given below.

4.4.7.1 Feminine plural of passive past participle aceso of verb acender

Input: /aseso/ i.e. stem ɛ[8] as per stem selection rule 144 (see worked example in 4.4.6.4).

The rule from Block 1 of the realisation rules that applies is:

170. RR_1,{vform: ppl, agr: {gend: fem},**V** ∩ **Pt**(<χ/o/, σ>) $=_{def}$ <[χ_i/a/], σ> for all i

This replaces the final /o/ of the input with /a/, giving the output /asesa/.

The rule from Block 2 of the realisation rules that applies is:

179. RR_2,{vform: ppl, agr: {num: pl},**V** ∩ **Pt**(<χ, σ>) $=_{def}$ <[χ_i/s/], σ> for all i

This concatenates the material /s/ to the input, giving us the final form /asesas/, as desired (orthographic *acesas*).

4.4.8 Capturing syncretism between the two past participles

In standard PFM1, syncretism is captured by rules of referral. This is also the case for Taylor's (2012) extension of PFM, as can be seen in the PFM rules adapted from her approach (Appendix III), where correspondences between whole cells in the paradigms are captured in 'Whole word syncretism rules', and correspondences between partial cells are captured in both stem selection rules and realisation rules by rules of referral using "σ/ϱ" notation.[73]

However, the apparent syncretism between the active and passive past participle forms is different from the examples of syncretism formalised by Taylor as rules of referral, in that it has exceptions,

---

[73] This notation represents that the features in the set σ are changed to include all the properties in the set ϱ. Taylor (2012, 372) quotes the following formal definition from Stump (2001, 56):

"Where σ and ϱ are well-formed sets of morphosyntactic properties, σ/ϱ is the smallest set such that:

a. for any atom-valued feature F:
   i. if ϱ is an extension of {F : v}, then σ/ϱ is an extension of {F : v}, and
   ii. if σ is an extension of {F : v} but there is no v' such that ϱ is an extension of {F : v'}, then σ/ϱ is an extension of {F : v};
b. for any atom-valued feature F:
   i. if ϱ is an extension of {F : v} and σ is an extension of {F : v'}, then σ/ϱ is an extension of {F : v'/v},
   ii. if ϱ is an extension of {F : v} and there is no v' such that σ is an extension of {F : v'}, then σ/ϱ is an extension of {F : v},
   iii. if σ is an extension of {F : v'} and there is no v such that ϱ is an extension of {F : v}, then σ/ϱ is an extension of {F : v'}."

namely the double past participle verbs. This class of exceptions means that rules of referral cannot account for the pattern; they would overgeneralise in producing a syncretism that exists for all verbs.[74]

As it happens, PFM can capture this pattern quite simply, without the need for additional rules of referral, essentially by not specifying a difference between two forms, making them identical by default.

This can be seen in the structure of the relevant complex stem selection rules:

143. SSR_0,{vform: ppl},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemc [10]], σ>

144. SSR_0,{vform: ppl}, (**IrregPTCP**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemɛ [8]], σ>

145. SSR_0,{vform: ppl, voice: active}, (**DoublePTCP**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemc [10]], σ>

By default, the complex stem selected as the active or passive past participle is the same, for both verbs with only a regular past participle form (by default, verbs in **V** ∩ **Pt**) and for verbs with an irregular past participle form (**IrregPTCP**). This implicitly captures the normal syncretism between the two functions, as the stem selection rules will give the same form ([stemc[10]]) for both active and passive past participles. Only in the class of **DoublePTCP** verbs is there a distinction. By default, the passive form will be [stemɛ [8]] (according to rule 144), and rule 145 specifies that the active form will be [stemc [10]].

[74] If a 'rules of referral' approach were taken further, additional rules could be written to specify the non-application of the rule of referral for the class of **DoublePTCP** verbs, but this seems unnecessarily complex, given the comparatively simple alternative solution presented here.

Under this analysis, there is no formal specification of syncretism within the PFM rules. The fact that the two past participle functions (active and passive) are by default filled by the same form is captured by the normal application and interaction of PFM rules.

Alternative analyses of the double past participles, with a different approach to capturing the syncretism, are of course available. In PFM2, inflection classes are treated as classes of stems, rather than of lexemes. This allows an account of 'heteroclisis', that is, lexemes whose paradigm includes stems of two different inflection classes. In this framework, we can treat the double past participle verbs as showing heteroclisis, with the regular and irregular past participle forms (stems) belonging to different inflection classes, while in most verbs, the irregular stem and the regular stem will be identical (syncretic).[75] In my analysis, the **IrregPTCP** and **DoublePTCP** inflection classes affect only the realisation of the past participle forms, which might be seen as undesirable, because we have to specify inflection class membership in relation to the whole lexeme in PFM1, even where it is not relevant to the vast majority of forms in the paradigm. The PFM2 analysis sketched here avoids this potential redundancy. However, there are advantages in being able to conceptualise a class of lexemes (not just stems) that participate in the double past participle phenomenon or have an irregular past participle: being able to refer to the **DoublePTCP** and **IrregPTCP** classes is useful in accounting for diachronic changes in these classes, as I demonstrate in Chapter 5.

Having considered various aspects of the morphology of the regular and irregular past participle forms, and the double past participle phenomenon, in the next section I consider the use of past participle forms in periphrases, and how the morphology of periphrases is treated in LFG-PFM.

---

[75] I am grateful to an anonymous reviewer for the International Symposium of Morphology (ISMo) for this suggestion.

4.4.9 Accounting for past participles in periphrases

The PFM rules laid out to account for past participles treat them as independent forms in the verbal paradigm, whether formed through pure inflection (in the case of active past participles) or as transpositions (passive past participles). However, as shown in section 4.1.1, a crucial function of both active and passive past participles is in periphrases.

Accounting for periphrasis in LFG has received considerable attention, as noted by Dalrymple, Lowe and Mycock (2019, 470). Various different approaches have been suggested, such as by Sadler and Spencer (2001), Ackerman and Stump (2004) and Biswas (2017). Some details of Biswas's proposals are discussed in section 3.3.4 above. Morphologists have also studied periphrasis from a PFM perspective (e.g. Spencer and Popova 2015), sometimes in combination with HPSG (Popova and Spencer 2012; Bonami and Samvelian 2015).

Bonami and Samvelian (2015) argue that not all the constructions that are 'typologically' identified as periphrasis should be analysed in the same way, where a syntactic construction fills a cell in an inflectional paradigm. Instead, they show that some Persian constructions normally identified as periphrases are instead better analysed as purely syntactic or purely morphological, when the data are examined closely. They also recognise that someone using a different approach might prefer to treat all these constructions as periphrases, for the sake of a more uniform analysis. It can be seen from their article that periphrasis is a complex subject which can be effectively treated in various different ways, depending on the assumptions made about the defining features of periphrasis, which vary from linguist to linguist, as well as between theories.

The approach given in this chapter to the morphology of past participles is in principle compatible with various different models of periphrasis. For Portuguese, given the clear parallelism of form between the passive periphrastic forms, which realise a particular value of the voice feature, and the active periphrastic forms, which realise tense/aspect properties, a morphological analysis that

treats the auxiliary verb and the past participle form together as realising a paradigm cell seems desirable, and I develop this analysis further in the rest of this section.

This analysis follows a common view of periphrasis in PFM (similar to the 'truly' periphrastic type identified by Bonami and Samvelian (2015)), where the periphrastic form fills a slot in the verbal paradigm to realise a set of morphosyntactic properties that otherwise might be realised through a one-word lexical item (a 'simplex form'). Spencer treats periphrasis as an example of 'a mismatch between morphological form and syntactic function' (2013, 161), in line with my LFG architecture approach in which morphological and syntactic properties are independent but linked. Spencer demonstrates how periphrastic forms such as the English progressive and perfect can be captured in a GPFM approach (2013, 221–22).

> '[The] English progressive and perfect aspects … are realized by a morphosyntactic construction which combines separate auxiliary verbs with specific morphological forms of the following verb. If the aspectual forms were synthetic, we would define a set of realization rules which specified the morphology of the word form realizing the [ASPECT:{progressive, perfect}] forms. For English, we could argue that the construction is still defined by a set of realization rules but ones which specify a periphrastic realization. Thus, the rules would state that [ASPECT:perfect], say, is realized by a syntactic template which includes the *have*-auxiliary followed by the perfect participle of the next verb in the syntactic sequence (for concrete proposals for Bulgarian periphrases see …[(Popova and Spencer 2012)], and for Persian periphrases see [(Bonami and Samvelian 2015)]). Part of the definition of the periphrasis would then be the specification of the content paradigm for the two components of the periphrasis. In principle, depending on the details of the construction, it would be open to us to define the verb forms purely in terms of m-features such as [$_{m}$ VFORM:{ing, en}] (or whatever).'

The principles he sets out are similar to those implemented by Biswas (2017) in his combination of LFG with PFM2. In this type of model of periphrasis, the past participle must exist as a cell in the paradigm, alongside the periphrastic form, which is compatible with my model of the paradigm, as shown in Figure 5, repeated here as Figure 8.

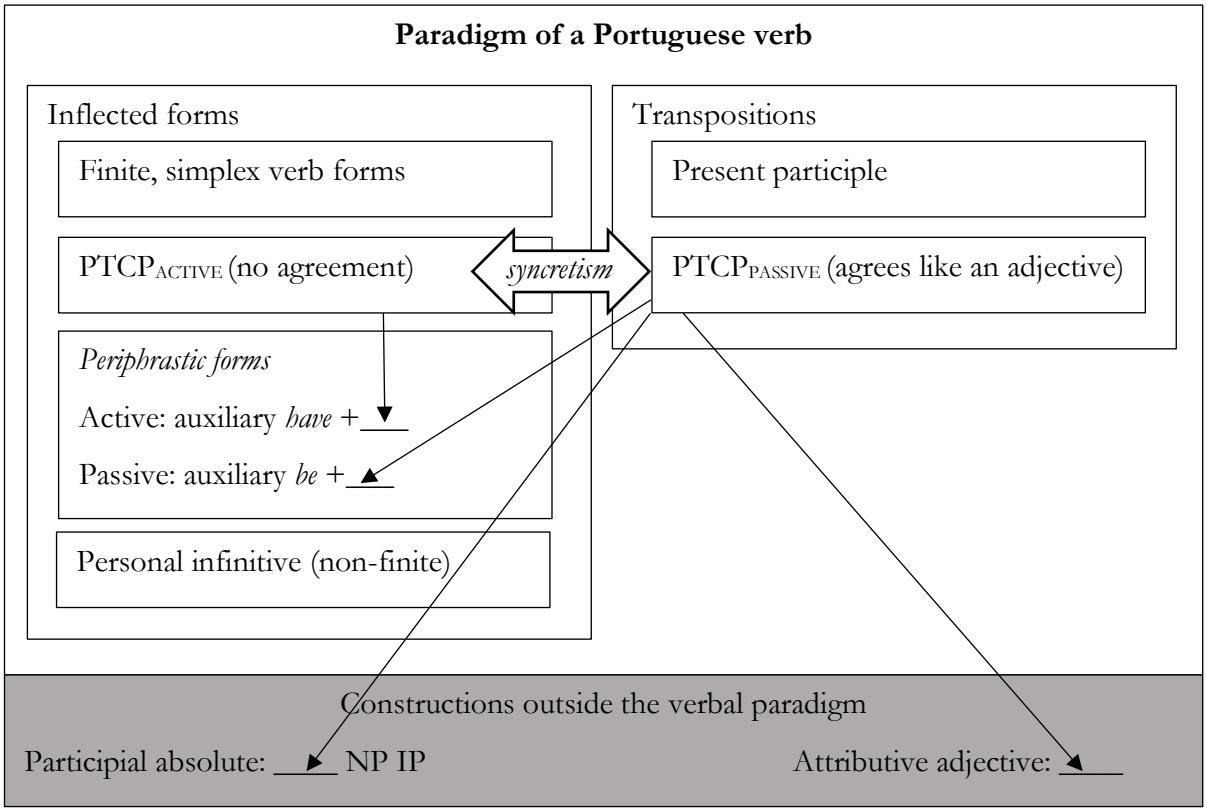


Figure 8 Past participles in the verbal paradigm

In section 3.3.4 I described Biswas's (2017) approach to periphrasis, based on PFM2. Here I adapt this approach in order to account for Portuguese periphrastic forms, using PFM rather than PFM2. The paradigm cells for periphrastic forms (box labelled '*Periphrastic forms*' in Figure 8) specify properties of the auxiliary form, and are also associated with collocation constraints that specify properties of past participles (which exist as separate cells). This approach is described here, using the examples of the periphrastic (active) pluperfect and the present passive.

Since there is a simple pluperfect in Portuguese, abbreviated 'plup' in Taylor's rules, and these new rules are for the compound (periphrastic) pluperfect, I refer to the compound pluperfect with the value for the tense feature 'plup 2' (on the model of Taylor's past 1 and past 2).

For the compound pluperfect with *ter* 'have', e.g. *tinha comido* 'I had eaten':

(135) If tns: plup 2, md: ind ∈ σ, PF<LI, σ> $=_{def}$ < PF<LI, (σ/ vform:ppl −{ tns:_ md:_})>, ω>

(136) If tns: plup 2, md: ind ∈ σ, ω = <{m-aux-c:+, m-lid-c:TER, m-tns-c:ipfv, m-mood-c:ind, m-dep-aux:-, m-dep-vform:ppl, m-dep-voice:active}, σ U {subjpers:_, subjnum:_}>

The rule numbered here as (135) can be paraphrased: if σ (the set of features to be realised) contains the pluperfect 2 (compound pluperfect) tense, and the indicative mood, then the realisation of the form has the same set of features σ with the additional feature vform: ppl (past participle), and with no tense or mood feature, and there is in addition the collocational constraint ω.

Similarly, rule (136) states that if σ (the set of features to be realised) contains the pluperfect 2 (compound pluperfect) tense, and the indicative mood, then the collocational constraint (ω) is that there must be an auxiliary, there must be a form of the auxiliary *ter* 'have', its tense must be imperfect, its mood must be indicative; and there must be a dependent which is not an auxiliary, with the vform: ppl feature, and its voice must be active; and the σ (feature set) of the auxiliary must contain the person value and number value of the subject.

For the present passive with *ser* 'be', e.g. *é comido* 'it is eaten':

(137) If tns: pres, md: ind, voice: passive ∈ σ, PF<LI, σ> $=_{def}$ < PF<LI, (σ/ vform:ppl −{ tns:_ md:_}) >, ω>

(138) If tns: pres, md: ind, voice: passive ∈ σ, ω = <{m-aux-c:+, m-lid-c:SER, m-tns-c:pres, m-mood-c:ind, m-dep-aux:-, m-dep-vform:ppl, m-dep-voice: passive, m-dep-gend:subj-gend, m-dep-num:subj-num}, σ U {subjpers:_, subjnum:_}>

Rule (137) states that if σ (the set of features to be realised) contains the present tense, the indicative mood, and the passive voice, then the realisation of the form has the same set of features σ with the additional vform: ppl feature and no tense or mood feature, and there is a collocational constraint ω.

Rule (138) states that if σ (the set of features to be realised) contains the present tense, the indicative mood, and the passive voice, then the collocational constraint ω is that there must be an auxiliary, there must be a form of the auxiliary *ser* 'be', its tense must be present, its mood must be indicative; and there must be a dependent which is not an auxiliary, with the vform: ppl feature, and its voice must be passive, and its gender and number must be the same of the subject; and the σ (feature set) of the auxiliary must contain the person value and number value of the subject.

These rules act as instructions for the realisation of the paradigm cell, describing the features of both the forms involved in the periphrasis, i.e. the auxiliary verbs (identified by their lexical identity or LID) and their dependents, the past participle forms. These rules for periphrasis contribute towards the c-structure realisation of the periphrastic forms, working alongside my PFM rules which ensure that the correctly inflected active and passive past participle forms, with their features, are available for the syntactic components of the grammar. However, the morphological rules that I have set out in this chapter are also compatible with alternative models of periphrasis, and I do not take a particular view on the correct syntactic analysis of periphrastic forms.

In the next section, I briefly demonstrate how this account of the morphological properties of periphrasis can stand together with an LFG analysis of further aspects of the passive, to capture the full range of grammatical behaviours across the different modules of the grammar: morphological, syntactic and semantic. This points towards the usefulness of the integration of PFM in the LFG framework.

4.4.10 An integrated LFG-PFM account of the passive periphrasis

The account of passive periphrases that I have developed so far here concentrates on their morphological properties, because my main focus is on the inflectional properties of the past participles used in the formation of the passive. The PFM rules given ensure that the correct form of the past participle (including agreement for number and gender) is selected and the collocation constraints ensure that the correct form of the auxiliary verb is selected. However, the passive has many other important and interesting properties beyond its morphological form. LFG is an ideal framework in which to consider the syntax and semantics of the passive, alongside the morphology. Although I do not attempt to provide a full account of these properties of the Portuguese passive, in this section I present a recent LFG account of the passive in English (Findlay 2016), using Glue Semantics and templates, which is eminently transferrable to Portuguese, to demonstrate how such an account can capture important aspects of the passive beyond its morphology.

First, a word about my treatment of voice. In the analysis given in section 4.4, voice (active/passive) is an inflectional morphological property. This morphological feature must also map to a syntactic VOICE feature. However, it is questionable whether VOICE should be treated as inflectional, within the morphological component. For example, in Spencer's (2013) GPFM framework, voice alternations are considered to be transpositions, because they have different FORM and SYN properties from the base lexeme but retain the same SEM and LI properties. Since my account of Portuguese morphology does not implement GPFM, I treat all periphrases, both passive and active, as if they were inflectionally related to the base lexeme (in line with my representation of the past participle, which is also a transposition: see section 4.4.2).

Using PFM as described throughout section 4.4, we have an account of the morphological form of the passive. In this section, I consider semantic aspects of the passive: in particular, how the passive alternation affects the alignment of grammatical functions with argument structure

positions or thematic roles. There are many different proposals for how to capture this alignment in LFG (Zaenen, Maling, and Thráinsson 1985; Levin 1986; Bresnan and Kanerva 1989; Kibort 2009; 2007; Findlay 2016). An overview of existing proposals and further information on the issues at stake is given by Dalrymple, Lowe and Mycock (2019, chap. 9). Here I sketch a recent proposal for analysing the passive implementing Kibort-Findlay Mapping Theory (Findlay 2016), a state-of-the-art approach to the mapping of grammatical functions.

Kibort-Findlay Mapping Theory aims to capture the different possible mappings between grammatical functions (SUBJECT, OBJECT, OBJECT_θ and OBLIQUE_θ)[76] and arguments (agent-like, patient-like, etc.). This theory of mapping has several important qualities, some of which follow from earlier approaches to mapping and related phenomena in LFG. Some general principles of the theory are summarised below, from more general to more specific ones.

- Argument structure is not treated as a separate module within the architecture. This allows a simple, direct mapping between c-structure and f-structure (following Asudeh and Giorgolo 2012)
- The proposals use Glue Semantics, and are therefore compatible with a resource-sensitive view of semantics (section 3.2.1)
- Semantic roles (agent, patient, goal, theme) are not treated as primitives, due to the lack of a satisfactory account of the range of thematic roles and their hierarchy (Kibort 2007)
- Functional description and other information is moved out of lexical entries and into templates, which are shared across many lexical items, again following Asudeh and Giorgolo (2012). (See discussion of templates in sections 3.3.1 and 4.4.2.)

[76] The subscript theta represents having a semantically restricted thematic role, e.g. goal.

- The cross-classification of grammatical functions used in standard Lexical Mapping Theory (Bresnan and Kanerva 1989), with the features ±r (restricted) and ±o (object-like),[77] is recast with four features MINUSO, PLUSO, MINUSR, PLUSR, defined as {SUBJ | OBL_θ}, {OBJ | OBJ_θ}, {SUBJ | OBJ}, {OBL_θ | OBJ_θ}

Findlay (2016, 326–27) gives an account of the English passive in this mapping theory. This account relies on several templates: one for transitive verbs (AGENT-PATIENT-VERB), one for the passive (PASSIVE), and one which generates a functional description to map a particular feature decomposition to a particular argument (MAP). I start by presenting MAP, as it is used in the definition of the other templates.

General MAP templates used for English are:

(139) MAP(MINUSO, ARG_1)

(140) MAP(MINUSR, ARG_2)

(141) MAP(PLUSO, ARG_3)

(142) MAP(MINUSO, ARG_4)

The full default mapping for each argument position contains a disjunction. This is demonstrated for ARG1 in (143) and ARG2 in (144). The second half of the disjunction in (143) states that nothing maps to ($\uparrow_\sigma ARG_1$), such that the mapping can be read: either a grammatical function that is MINUSO maps to ARG_1, or nothing maps to ARG_1:

(143) {@MAP(MINUSO, ARG_1) | ($\uparrow_\sigma ARG_1)_\sigma^{-1}$=Ø}

(144) {@MAP(MINUSR, ARG_2) | ($\uparrow_\sigma ARG_2)_\sigma^{-1}$=Ø}

[77] The properties of the grammatical functions are as follows: SUBJECT is −r, −o, OBJECT is −r, +o, $OBJECT_\Theta$ is +r, +o and $OBLIQUE_\Theta$ is +r, −o.

The final forms of these templates abbreviate the second half of disjunction using a template @NOMAP:

(145) ARG1 := {@MAP(MINUSO, ARG_1) | @NOMAP(ARG_1)

(146) ARG1 := {@MAP(MINUSR, ARG_2) | @NOMAP(ARG_2)

These templates capture the typical alignment of the argument positions (ARG_1 etc.) with the feature specifications (MINUSO etc.).

An additional mapping used for the passive is:

(147) MAP(PLUSR, ARG_1)

This adds PLUSR (equivalent to +r) to the specification for ARG1: in other words, it captures the insight that in the passive, the normal ARG1 is demoted from subjecthood to have the features +r, −o (i.e. it is $OBLIQUE_\theta$).

The AGENT-PATIENT-VERB valency template (for transitive verbs) is also used to account for the passive. In Findlay's presentation, this calls the templates @ARG1 and @ARG2, as well as providing semantic information:

(148) AGENT-PATIENT-VERB :=

@ARG1

@ARG2

$\lambda P \lambda y \lambda x \lambda e.P(e) \wedge agent(e) = x \wedge patient(e) = y$:

$[(\uparrow_\sigma \text{EVENT}) \rightarrow \uparrow_\sigma] \rightarrow (\uparrow_\sigma \text{ARG}_2) \rightarrow (\uparrow_\sigma \text{ARG}_1) \rightarrow (\uparrow_\sigma \text{EVENT}) \rightarrow \uparrow\sigma$

There is also a template for the passive itself:

(149) PASSIVE ≔

(↑VOICE) = PASSIVE

{@MAP(PLUSR, ARG_1) | @NOMAP(ARG_1)}

($\lambda P.\exists x[P(x)]$: [(↑$_\sigma$ ARG_1) → ↑$_\sigma$] → ↑$_\sigma$)

The line '(↑VOICE) = PASSIVE' supplies the VOICE attribute with the feature PASSIVE for f-structure. Crucially for my purposes, this provides the morphosyntactic property (VOICE: PASSIVE) that allows the selection of the passive periphrasis within the morphological component, for the correct realisation of this form.

The second line accounts for the 'demotional' property of the passive, i.e. it provides the +r value for ARG1, ultimately meaning that this is realised as $OBLIQUE_\theta$ rather than SUBJECT.

The third line 'existentially closes a dependency on the meaning of ARG1' (Findlay 2016, 326), which is to say that this provides the appropriate meaning for ARG_1 in the short passive, i.e. when there is no agent by-phrase.

When all these elements are put together, to resolve the mapping disjunctions, ARG_2 must be realised as SUBJECT and ARG_1 may either be $OBLIQUE_\theta$ or may be omitted.

This approach allows a full account of the semantic aspects of the active/passive alternation, which can be referred to in a simple way in the lexical entry, using templates. It provides a unified approach to the passive that accounts for the changes in alignment between argument positions and grammatical functions, without necessitating a further module within the LFG architecture (a-structure), as required in some accounts. Given the similarities between the English and Portuguese passives, this approach could be directly transferred to Portuguese to account for the active-passive alternation, alongside the morphological account of periphrases presented in section 4.4.9.

Throughout section 4.4 I have shown how various aspects of the formal LFG-PFM model are combined to give an accurate and precise analysis of Portuguese double past participle forms. In the next section, I bring together some conclusions from the whole of this chapter.

4.5 Conclusions

This chapter has shown the power of PFM, in combination with LFG, to give an effective analysis of the double past participles in Portuguese, a complex phenomenon that involves both syntax and morphology. Both theories are required to capture different aspects of the phenomenon. In the analysis presented here, the PFM-based morphological component fulfils the need to form the correct inflected forms of the past participle according to their function, while the LFG lexical/lexemic entries store crucial information about the forms. Finally, we have seen that the morphological side of the analysis exists alongside a powerful LFG account of the semantic and syntactic aspects of the active-passive distinction, which is highly relevant to the past participles' use in periphrastic forms.

My model captures the differences in morphological agreement between active and passive forms at the level of realisation rules, while capturing the similarities between the forms through the fact that the forms are by default the same (complex) stem. This analysis draws on both Taylor's (2012) adaptation and Spencer's (2013) extension of PFM to give a new and detailed account of Portuguese double past participles. My account separates out the different forms and functions of past participles, which have not previously been captured in a PFM analysis. I also demonstrate how applying Spencer's theory of lexical relatedness beyond pure inflection and derivation captures the adjectival qualities of Portuguese passive past participles, which can be analysed as transpositions. This reflects the differences between different past participle functions in terms of morphosyntactic agreement, internal syntax and distribution. By combining PFM-based tools for morphological analysis with standard LFG assumptions about the structure of the grammar, my

model can accurately account for the occurrence of past participles with the right properties in the right contexts.

Double past participles in European Portuguese thus represent a useful case study for the combination of PFM and LFG to account for complex morphosyntactic phenomena. The proposals made here for a more detailed treatment in these formal models of non-finite verbal forms, and the periphrases which contain them, also have wide-reaching applications in other languages, both within the Romance family and beyond. These forms have received perhaps less attention than deserved in previous LFG and PFM literature. My work contributes to filling this gap by presenting an analysis that recognises and accounts for the complexities involved.

This chapter has dealt only with the synchronic patterns of distribution of forms between functions found in European Portuguese. However, the diachronic trajectories of the double past participle forms merit closer consideration. For example, some verbs in the second and third conjugations developed new regular past participles and later lost their irregular past participles. On the other hand, some verbs in the first conjugation that used to have only a regular past participle now have only an irregular past participle. Members of both of these classes of verbs passed through an intermediate stage where they exemplified the double past participle phenomenon. Other verbs show a more stable membership of the double past participle inflection class. The next chapter investigates how these trajectories can be described by the formal theoretical model presented here, giving new insights into the patterns seen in the historical data and suggesting further applications of LFG-PFM.

Chapter 5 Portuguese double past participles in diachrony

This chapter presents the development of double past participle forms in the history of Portuguese, including the divergence of different varieties, and provides rules to account for this in the formal PFM-LFG framework.

PFM and LFG have both been developed to account for synchronic linguistic phenomena, focusing on the rules required to account for language patterns at one point in time. I argue that these theories can also be applied fruitfully to diachronic phenomena, continuing recent lines of research in LFG: see for example Vincent (2001), Börjars et al (2016), and discussion in section 3.2.2. This chapter focuses on showing how PFM analyses can be adapted to account for diachronic data, and (in section 5.3.5) assesses how these changes are reflected at the level of the LFG lexemic entry, showing the power of the combined LFG-PFM approach. I also demonstrate in Chapter 6 how other parts of the LFG architecture, in combination with PFM, can reflect different types of grammatical change associated with grammaticalised pronouns.

Language change can be captured in PFM through making several separate static analyses, which reflect diachronic change by showing the grammar of a language as used by a speaker or a group of speakers at distinct moments. Through reanalysis, a later speaker (e.g. a member of the next generation) may come to have a grammar with one or more incremental changes, represented by slightly different rules. In fact, the same speaker's mental representation of a grammar may also change over time: there is 'a growing body of evidence that linguistic change is continuous throughout the lifetime of an individual' (Blevins and Blevins 2009, 5).

This chapter builds on Chapter 4 by providing a diachronic analysis of the double past participle phenomenon in Portuguese. To this end, I begin by setting out how the double past participle phenomenon in Portuguese has changed over time, presenting new data extracted from historical corpora of Portuguese. I trace the Latin origins of modern Portuguese verbs' past participles,

through their historical development in Portuguese, and consider divergences in the double past participle phenomenon in the present day between European and Brazilian Portuguese.

In the case of the double past participles, most of the differences between the representations occur at the level of the lexeme, in the lexemic entry.[78]

The diachronic trajectory of each verb is different. Verbs may be reanalysed into different inflection classes, meaning that different rules apply to them. For example, a verb that used to be in the class of verbs with two past participle forms (represented in my PFM rules as the **DoublePTCP** class) might leave this class, while remaining in the **IrregPTCP** class of verbs with irregular past participles. Varieties may also diverge in terms of the PFM rules that describe their morphological systems. For example, the distribution of regular and irregular past participle forms between active and passive contexts may vary.

The chapter is structured as follows. Section 5.1 sets out the diachronic trajectories of Portuguese verbs which have (or had) two past participle forms. As with any diachronic study, some care must be taken when considering the evidence, so the section begins with a brief methodological note (section 5.1.1). I then show that from Latin (section 5.1.2) and throughout the history of Portuguese (section 5.1.3), there is considerable variation in the paths of development of the past participle forms. Corpus data regarding the frequencies of different past participle forms are summarised in section 5.1.4. Section 5.1.5 presents information about Brazilian Portuguese, where we see some divergence compared with European Portuguese.

Section 5.2 begins the process of analysing these changes from a formal perspective, first considering the differences between the major inflection classes (section 5.2.1), before looking at

[78] Sample lexemic entries are given in section 5.3.5. As we saw in Chapter 4, some of this information can be represented in the table format of the lexical data given in Appendix III.

the role of lexical specification (section 5.2.2) and analogical change (section 5.2.3) in the LFG-PFM model.

Section 5.3 presents my LFG-PFM analysis of the different diachronic trajectories taken by verbs in the history of Portuguese up to modern European and Brazilian varieties, in terms of the existence of a regular and/or irregular past participle. Section 5.3.1 considers the role of stem roots in diachronic change of these past participles, and section 5.3.2 looks at the interaction of inflection classes and stem formation rules in accounting for different past participle forms. Section 5.3.3 presents more detailed case studies of developments seen in some Brazilian Portuguese varieties, namely the innovative past participles and the loss of the active/passive distinction. Section 5.3.4 briefly considers how the cross-paradigm phenomenon of reanalysis of paradigm membership might be captured in the PFM model, and finally, section 5.3.5 indicates how morphological change is represented at the level of the lexemic entry in LFG-PFM.

This chapter demonstrates the effectiveness of the LFG-PFM model in capturing diachronic developments in morphological phenomena at both fine-grained and broader-scale levels of analysis. Conclusions from the chapter are presented in section 5.4.

5.1 Diachronic trajectories of double past participle forms

There is considerable variation in the diachronic trajectories of Portuguese verbs that show the double past participle phenomenon. This section describes the diachronic development of the past participle forms of these verbs, from the existence of alternative forms in Latin for some verbs to the innovation of regular and irregular forms in the history of Portuguese.

The section also considers the difficulties of reconstructing the usage of the different forms at different points in time, due to the nature of the evidence available. In particular, it is not always possible to determine restrictions on the contexts in which different past participle forms could occur, although some generalisations can be drawn. In Chapter 4 we briefly saw differences in the

past participle forms between European Portuguese and Brazilian Portuguese. This section gives context, supplemented by further detail in section 5.2, for a formal theoretical analysis of the historical development of the different forms, in section 5.3.

It is crucial to trace the development of the irregular and regular past participle forms separately. Lobato (1999), following Said Ali ([1908] 1957), argues that for verbs with distinct forms, the two forms are of separate origin: the short (irregular) form is not derived from the long (regular) form, nor is the long form derived from the short form. Lobato emphasises that according to historical evidence, there is no necessary chronological order of the formation of the two past participle forms.

The occurrence of two distinct past participle forms in a single paradigm can develop in a variety of different ways, as described in sections 5.1.3 to 5.1.5. There are instances when the past participles of two different Latin verbs have been merged into one paradigm within Portuguese, as well as instances where a new form is innovated in Portuguese. A recurring theme throughout this chapter is that the history of past participle forms is not simple, and no single explanation exists for all the different development patterns.

5.1.1 Methodological questions in diachronic corpus work

This section discusses some of the difficulties in tracing the history of the Portuguese double past participle forms and their predecessors in Latin. There is only a relatively small amount of evidence for these forms, as many of the verbs are not frequently attested in the surviving written evidence.

In historical work, we have no access to individual speakers of historical varieties to ask about their intuitions, so corpora provide the only evidence available to us. As well as the usual problems involved in studying modern corpora (such as combining the usage patterns of many different people, which can disguise underlying patterns), historical corpora have their own limitations. The survival of Latin texts and historical Portuguese texts is essentially random, with no guarantee of

a particular level of representation of any given verb or verb form. Even when looking at fairly well attested time periods, we may have zero or extremely few examples of the use of a particular verb in the past participle form. Extrapolating the patterns of usage, such as the syntactic contexts in which particular verb forms occur, is often impossible, so we cannot have a completely clear or accurate picture of how forms were distributed between different functions at any point in time. Dating the first occurrence of particular forms is also problematic, given the patchy nature of the written record. Innovative vernacular forms may not be attested, due to conservative or Classical linguistic norms.

The Latin corpus used for evidence in this work is PHI Latin Texts (Packard Humanities Institute 2015), an open-access corpus which includes 'essentially all Latin literary texts written before 200 AD, as well as some texts selected from later antiquity', providing an excellent source of Latin data. Although, as noted in section 1.3, Portuguese ultimately descends from spoken rather than literary Latin, the literary language provides a good proxy in the absence of comprehensive written evidence of the everyday spoken language.

The Portuguese corpora used are the *Dicionário de Verbos do Português Medieval*, which lists every wordform from the *Corpus Informatizado do Português Medieval*, and the *Corpus do Português: Genre/Historical* (Davies and Ferreira 2006), which contains 45 million words of Portuguese from the 14th to 20th centuries. Together these corpora give a large quantity of good quality evidence, allowing clear insight into the historical development of Portuguese.

With the general limitations of historical corpus work in mind, the picture given here of diachronic developments from Latin and throughout the history of Portuguese is necessarily limited, aiming to give a sense of these developments without claiming to put forward a definitive history of all the verbs and individual verb forms involved, which would be a much greater task. I do not make any strong claims regarding the dates at which particular forms were introduced, or the syntactic

contexts that forms were used in at different points in time, but instead demonstrate the pathways of development seen in the past participle forms of a sample of Portuguese verbs.

5.1.2 Past participles in Latin

Many Portuguese past participles are regular reflexes of Latin perfect passive participles, which are characterised as verbal adjectives: they were adjectives formed from a verbal root, which could occur as part of the verb paradigm in passive periphrases. Portuguese forms continue to be used in some contexts in which Latin perfect passive participles also occurred. However, different verbs show different patterns in terms of the extent to which their past participle forms are derived from Latin forms or innovated in Portuguese.

To understand the context of these developments, I briefly set out here the relationship between Latin and Portuguese conjugation classes (section 5.1.2.1), and the typical patterns of past participle forms in the different classes in Latin (section 5.1.2.2).

5.1.2.1 Overview of Latin and Portuguese conjugation classes

A general schema of the numbering of Latin verb conjugations and their relationship to Portuguese conjugations is given in Table 21, using facts from Stovicek (2010, 88) who discusses the development of the conjugation classes across Hispano-Romance from Latin. Stovicek's account is based on Redenbarger's (1976) five-conjugation system for Latin. In this system, the classes are characterised by a thematic vowel, except class III, which is athematic.[79]

[79] Verbs are given in the infinitive and in the first person singular present indicative. Where relevant, long vowels are represented with a macron (e.g. *ā*) and short vowels with a breve (e.g. *ĭ*).

Latin	I - *ā*	II - *ē*	III - ∅	IIIi - ĭ	IV - *ī*
e.g.	*laudāre* 'praise'	*monēre* 'warn'	*dūcere* 'lead'	*capĕre* 'take'	*audīre* 'hear'
	laudō 'I praise'	*moneō* 'I warn'	*dūcō* 'I lead'	*capiō* 'I take'	*audiō* 'I hear'
Portuguese	1 *-ar*	2 *-er*	*[2 -er]*	*[2 -er]*	3 *-ir*
e.g.	*cantar* 'sing'	*comer* 'eat'			*partir* 'break'
	canto 'I sing'	*como* 'I eat'			*parto* 'I break'

Table 21 Development of Portuguese verb conjugations from Latin

Classes III and IIIi are sometimes regarded as two sub-conjugations of a single conjugation, III. According to Stovicek, Class III verbs were the most numerous of classes II-IV in Classical Latin, while there were few Class IIIi verbs, and these verbs were mainly lost in the development of Hispano-Romance. Some Classical Latin verbs from conjugations II-IV survived in Hispano-Romance only as *-ar* verbs formed on the basis of their perfect passive participles, such as Latin *proicere* 'throw out', participle *proiectum*, which survives as Portuguese *projetar* 'project'.[80]

For further details on Latin conjugations, including the alternation of [ĕ] and [ĭ] in conjugations III and IIIi, I refer the reader to Stovicek (2010).

[80] Stovicek notes that a similar process is also seen in Portuguese *frigir* 'fry' whose irregular past participle *frito* has led to the development of a first conjugation verb *fritar*. Evidence from the *Corpus do Português* (Web/Dialects) suggests that *frigir* has now fallen out of use other than in the set phrase *no frigir dos ovos*, lit. 'when the eggs are fried' or 'at the end of the day'.

5.1.2.2 Latin perfect passive participle forms of verbs across conjugation classes

Nearly all Latin verbs have a perfect passive participle (PPP), normally considered a 'verbal adjective', which agrees in number, gender and case (like other Latin adjectives). I give examples here in the masculine singular form. The perfect passive participle follows a general pattern of:

(150) stem (+ TV) + t + *us*

The vowel that occurs before the *-tus* ending is the thematic vowel, in the case of conjugations I and IV, and for some verbs in conjugation II. Verbs in conjugation II may use *-ĭtus* instead of *-ētus*, such as *monitus* below.

It is also possible for these participles to be athematic if they have no vowel before the /t/. This is the regular pattern in conjugation III/IIIi, and some vowels in conjugation II also follow this pattern.

Class	I - *ā*	II - *ē*	III - ∅	IIIi - ĭ	IV - *ī*
Infinitive	*laudāre* 'sing'	*monēre* 'warn'	*dūcere* 'lose'	*capĕre* 'take'	*audīre* 'feel'
PPP	*laudatus*	*monitus*	*ductus*	*captus*	*auditus*

Table 22 Past participles across Latin conjugations

There are some variations from this basic pattern (Harrison n.d.). In some verbs, regular alternations apply, namely devoicing of the final consonant of the stem and/or lenition of the consonant /t/:

Sound change	**Outcome**	**Example**	**Conjugation class**
g + t > ct	'-ctus'	*tingo* 'dip' > *tinctus*	III
b + t > pt	'-ptus'	*scribo* 'write' > *scriptus*	III
d + t > ss (> s)	'-ssus' or '-sus'	*sedeo* 'sit' > *sessus*	II
t + t > ss (> s)	'-ssus' or '-sus'	*mitto* 'send' > *missus*	III

Table 23 Sound changes in Latin participles

Some of these verbs therefore have forms ending *-sus* due to phonological processes; other verbs also have forms of this type by analogy with those verbs.

According to Harrison (n.d.), some Latin verbs also had forms in *-utus*, where vocalic /u/ replaced consonantal /v/ in the stem, e.g. *solvo* 'untie' > *solutus*. This pattern later spread to other verbs in Portuguese, as well as in other Romance varieties, to give a series of past participles in *-udo*, which have later been supplanted by *-ido* in Portuguese (Laurent 1999, 305–9; Maiden 2011, 204), giving rise to the modern two-way distinction in regular Portuguese past participles between *-ado* and *-ido* forms.

Latin perfect passive participles can therefore be classified as either *thematic* or *athematic*, and the athematic forms can be further divided into *sigmatic* and *tautic* forms (Laurent 1999): sigmatic forms contain an /s/, coming from an earlier *-sus* ending, while tautic forms have the ending *-tus* suffixed directly to the stem (with no intervening thematic vowel).

The past participles of Portuguese verbs which directly continue from Latin perfect passive participles in classes II or III may either have thematic *-ido* or athematic *-o* endings. They may also continue /t/ endings or /s/ endings, as described in section 4.1.2.

We have seen in this section that Latin perfect passive participles existed in a variety of forms, which were not fully predictable from conjugation class. As we shall see in the next section, the development of participle forms from Latin to Portuguese has introduced further complexity.

5.1.3 Double past participle forms from Latin to Portuguese

Past participles in Portuguese may be inherited from Latin as reflexes of the Latin perfect passive participle forms described above, or they may be inherited from other Latin forms, such as simple adjectives. Alternatively, they may have been innovated at some stage within the history of Portuguese.

Several distinctive patterns of innovation are seen within Portuguese past participles. Second and third conjugation verbs in Portuguese were more likely to inherit athematic forms from Latin, and in some of these verbs, additional thematic forms were innovated within Portuguese.

On the other hand, first conjugation verbs in Portuguese typically inherited regular thematic forms from Latin. Some of these verbs also innovated athematic forms, probably by analogy with simple adjective forms. Other verbs incorporated existing simple adjectives into their paradigm as past participles. As we saw in Chapter 4, the distribution of simple adjectives and past participles overlaps considerably (such as in the 'copula + _' context), leading to a natural tendency for the two types of forms to become associated and for analogical processes to act between forms in the two contexts. In LFG, both adjectives and past participles in this context after a copula could be treated as predicates in the grammatical function PREDLINK, reflecting the parallelism between the two structures.[81]

In this section, I look in more depth at the trajectory of a sample of double past participle verbs, which together exemplify the main patterns of diachronic change in the phenomenon. For convenience, I divide these between Portuguese first conjugation verbs (section 5.1.3.1), and Portuguese second and third conjugation verbs (section 5.1.3.2).

[81] There are several different proposals for the analysis of copular constructions (with or without a copular verb) in LFG. I refer the reader to Dalrymple, Lowe and Mycock (2019, sec. 5.4.5) for details.

5.1.3.1 First conjugation verbs

For some verbs, both irregular and regular past participle forms can be traced back to Latin. An example of this is the Portuguese verb *aceitar* 'accept', whose two past participle forms *aceite/aceito*[82] and *aceitado* can both be traced to distinct Latin forms, *acceptus* and *acceptatus*.[83] *acceptus* was the past participle of *accipio* 'accept', while *acceptatus* was the past participle of the frequentative verb *accepto* 'keep accepting'. Latin frequentatives are formed on the basis of the supine (a 'verbal noun' whose stem is shared with other forms, including the past participle) of their base verb: in this case, the supine *acceptum* from *accipio* gives the frequentative *accepto*. The Portuguese verb *aceitar* is derived from the frequentative *accepto*, but incorporates the original past participle *acceptus* (*aceite/aceito*) as well as the past participle of the frequentative *acceptatus* (*aceitado*).

For other verbs, only one form can be traced back to Latin: this appears to be the case for *pagar* 'pay', as reported by Lobato and Said Ali. The historically attested regular Portuguese past participle *pagado* is derived from the regular Latin perfect passive participle *pacatus* from *paco* 'make peaceful'. There is no evidence of the existence of a Latin form like **pacus*. The irregular Portuguese past participle *pago* was likely formed by analogy with the nominal form *paga* 'payment' (Ali [1908] 2008, 159). Maiden (2013, 506) alternatively states that the model for the class of 'short' irregular past participles in first conjugation (*-ar*) verbs was 'the large class of disyllabic adjectives in *-o*'. The form *pagado* has later fallen out of use in Portuguese, leaving only the irregular *pago*.

---

[82] The form *aceite* is used in European Portuguese while *aceito* is used in Brazilian Portuguese. See footnote 104 (p258) for more information on this vowel difference.

[83] As shown in section 5.1.4, the attested history of this verb in Portuguese only starts at the 15th century, so it is impossible to say with certainty whether both forms were directly inherited from Latin, but both forms have an equally long attested history, starting in the 15th century.

The history of the Portuguese verb *gastar* 'spend' is different again. The Latin verb *vasto* 'lay waste' was originally derived from an adjective *vastus* 'empty'. From the verb *vasto* comes the regular past participle *vastatus* in Latin, giving the Portuguese regular past participle *gastado*, influenced by Germanic **wôstjan* (Porto Editora 2003c). The form *gastado* is attested in Portuguese from the 1400s to the 1700s in active contexts. The modern Portuguese short form *gasto*, which is the irregular past participle of *gastar* 'spend', may have come into existence by analogy with the form *pago*, as Ali ([1908] 2008, 159) claims. The long form *gastado* has subsequently fallen out of use in European Portuguese, leaving only the irregular *gasto*.

Another first conjugation verb with a similar history is *ganhar* 'earn', which has a complex etymology: its source is the Germanic verb **waidanjan* 'gather food, hunt', coming into Portuguese via the French *gagner* 'earn' (Porto Editora 2003b; Brosman 1963). The Portuguese verb had a regular past participle form *ganhado* which has subsequently been replaced by the irregular *ganho*. The forms *gasto* and *ganho* both existed as nouns in older forms of Portuguese (Ali [1908] 1957, 159), suggesting that they have both been reanalysed as past participles by analogy with *pago* possibly due to their similarity in meaning. In addition, it is possible that *ganhar* was influenced by *gastar* in this respect (see section 5.3.3).

The verb *limpar* 'clean' has a similar history to *gastar* 'spend', with the original Latin adjective *limpidus* 'clear' giving a Latin verb *limpido* with a reconstructed regular past participle **limpidatus* > Port. *limpado* (regular past participle). The original Latin adjective may also have been conserved as the Portuguese adjective *limpo* 'clean', which became an irregular past participle and is now the only past participle form of *limpar*.

As a final example of the many possible paths of development by which first conjugation verbs came to have two past participle forms, the Portuguese verb *matar* is originally from Latin *mactare* 'slaughter', giving the past participle *mactatus* > Port. *matado* (regular past participle). The modern

Portuguese verb paradigm also includes the suppletive irregular past participle *morto*, which is derived from Latin *mortuus*, the perfect active participle of *morior* 'die'.[84]

Generalising across these first conjugation verbs, then, the Latin verbs typically had regular past participle forms in *-atus* > modern Portuguese *-ado*. However, alongside these regular past participles, there were purely adjectival forms that had in some instances given rise to verbs, and thus shared some features with them.[85] The existence of a small number of these adjectives, and their close similarity to the true past participles (verbal adjectives), laid the path for their inclusion in the verb paradigm as irregular past participles in Portuguese. This in turn led to further irregular (athematic) past participles being innovated in other first conjugation verbs by analogy.

This analogical extension of athematic past participles is also seen in the more recent history of Portuguese, in the development of so-called innovative irregular past participles in Brazilian Portuguese (Chagas de Souza 2007; 2011; 2015; Nevins and Rodrigues 2014), discussed in section 4.3.3. This pattern is primarily seen in first conjugation verbs such as *pegar* 'take, grab', but has also extended to verbs in other conjugations such as *pedir* 'ask', as discussed in section 5.1.5 below. These innovative forms are identical to the first person singular present indicative of the relevant verb, which has been claimed to be an analogical development which has led to a synchronic syncretism (Chagas de Souza 2007; 2015; 2011).

---

[84] *morior* has no perfect passive participle, only a perfect active participle, because it is a deponent verb, active in form and passive in meaning.

[85] To use the terminology of Spencer (2013) these forms likely share at least the properties of FORM and SEM with the verb, since they share morphological and semantic properties, but not SYN (syntax) and LI (lexical index), which would make them a type of transpositional lexeme.

A summary of the different trajectories of some first conjugation verbs, including *pegar* 'take, grab' as an example of the Brazilian innovative pattern, is given in Table 24. In this table I bring together forms attested in Portuguese with the forms they relate to in Latin, as attested in the PHI Latin Corpus, and include forms that are likely to have been innovations within the history of Portuguese, according to evidence from the written language represented by historical corpora of Portuguese (see section 5.1.4). Forms which have been proposed to exist but are not attested in the PHI Latin Texts corpus (Packard Humanities Institute 2015) are marked with an asterisk. The citation form used for Latin verb lexemes is the first person singular present indicative.

Portuguese verb	**Latin verb**	**Latin form(s)**	**Innovated form**	**Modern forms**
aceitar 'accept'	*accipio* 'accept' *accepto* 'keep accepting'	*acceptus* **acceptatus*	-	*aceito* *(aceitado)*
entregar 'hand over'	*integro* 'renew'	*integer* (adj) *integratus*	-	*entregue* *entregado*
gastar 'spend'	*vasto* 'lay waste'	*vastus* (adj) 'empty' *vastatus*	-	*gasto* *(gastado)*
limpar 'clean'	**limpido* 'clean'	*limpidus* (adj) 'clear' *limpidatus*	-	*limpo* *(limpado)*
matar 'kill'	*macto* 'slaughter'	*mactatus* *mortuus* < *morior* 'die'	-	*matado* *morto*
pagar 'pay'	*paco* 'make peaceful'	*pacatus*	*pago*	*pago* *(pagado)*
pegar 'take, grab'	*pico* 'smear with pitch'	*picatus*	*pego*	*pego* (BP) *pegado*

Table 24 Trajectories of past participles of first conjugation verbs.

The Latin form *acceptatus*, which is not attested in the Latin corpus, can be reconstructed as the likely perfect passive participle forms of the frequentative verb *accepto* 'keep accepting'; similarly, *limpidare* is not attested in the corpus but may have existed in the Vulgar Latin spoken in Iberia, and is attested in Late Latin (McCarren and Ritter 2020).

5.1.3.2 Second and third conjugation verbs

Several verbs in the Portuguese second conjugation (*-er*) and third conjugation (*-ir*) that show or have shown the double past participle phenomenon have irregular forms corresponding to athematic Latin perfect passive participles. As we saw in the previous section, Latin verbs in the antecedents of the Portuguese second conjugation could have either athematic or thematic forms.

The Latin verbs *accendo* 'kindle', *eligo* 'choose', *suspendo* 'suspend' and *prendo* 'take' (variant of *prehendo*) have the participle forms *accensus, electus, suspensus* and *prensus*, which give the Portuguese irregular past participle forms *aceso*, *eleito*, *suspenso* and *preso*. The regular past participle forms of these verbs in modern Portuguese were derived later, by analogy with existing past participles in *-ido*, from Latin thematic forms: *acendido, elegido, suspendido, prendido*. Both the regular and irregular forms of these verbs survive into modern Portuguese, with the irregular forms broadly being retained in passive contexts and the regular forms in active contexts, at least in some varieties.

According to Lobato (1999), the Portuguese verbs *absolver* 'absolve' *absolvido/absolto* and *defender* 'defend' *defendido/defeso* are further examples from the Portuguese second conjugation where the regular form has been innovated while the irregular form was inherited. Another Portuguese verb, *cingir* 'gird' which has the forms *cingido* and *cinto*, from the third conjugation, shows the same trajectory. However, the survival of these irregular forms as past participles in Portuguese is extremely limited. Searches on the *Corpus do Português* (Web/Dialects) show that the infrequently used verb *cingir* retains the past participle form *cingido* while *cinto* exists only as a noun meaning 'belt'. The form *defeso* occurs only in the specialised adjectival meaning 'closed to entry' and the nominal meaning 'closed season (for fishing or hunting)', while *defendido* acts as the past participle

for *defender* 'defend' in all contexts. Finally, *absolto* never occurs in modern Portuguese, with *absolvido* used as the past participle of *absolver*.

The second and third conjugation verbs thus in general have a different diachronic trajectory of development from that seen in the first conjugation verbs. However, we should bear in mind that the patterns are limited to very small numbers of verbs in any conjugation. The vast majority of Portuguese verbs of all conjugations have only one (regular) past participle form. The implications of this are discussed further in section 5.2.

The different patterns found in these second and third conjugation Portuguese verbs are shown in Table 25, which brings together the same types of data as described above for Table 24.

Portuguese verb	**Latin verb**	**Latin form(s)**	**Innovated form**	**Modern forms**
absolver 'absolve'	*absolvo* 'loosen'	*absolutus*	*absolvido*	*(absolto)* *absolvido*
acender 'light/turn on'	*accendo* 'light'	*accensus*	*acendido*	*aceso* *acendido*
cingir 'gird'	*cingo* 'surround'	*cinctus*	*cingido*	*(cinto)* *cingido*
defender 'defend'	*defendo* 'drive away'	*defensus*	*defendido*	*(defeso)* *defendido*
eleger 'choose'	*eligo* 'choose'	*electus*	*elegido*	*eleito* *elegido*
prender 'arrest'	*prehendo* or *prendo* 'seize'	*prehensus* or *prensus*	*prendido*	*preso* *prendido*
suspender 'suspend'	*suspendo* 'hang up'	*suspensus*	*suspendido*	*suspenso* *suspendido*

Table 25 Trajectories of past participles of second and third conjugation verbs

As we have seen in the last two sections, there are various different patterns in the development of past participle forms in Portuguese from the corresponding perfect passive participles in Latin. Different types of forms have been inherited and innovated, with some tendencies across the major inflection classes (first, second and third conjugation) in the retention of these different forms. In section 5.1.4 I present further analysis of the forms, using corpus data from the history of Portuguese.

5.1.4 Attestations of different forms in the history of Portuguese

This section presents information on the attestations of different past participle forms for a set of verbs throughout the history of Portuguese. Ten transitive verbs were selected for study: six from the first conjugation (namely *aceitar* 'accept, *entregar* 'hand over', *gastar* 'spend', *limpar* 'clean', *matar* 'kill' and *pagar* 'pay'), and four from the second conjugation (namely *acender* 'light, turn on'*, eleger* 'choose', *prender* 'arrest' and *suspender* 'suspend'). The verbs were chosen as they are relatively high frequency lexical items, so they should be represented more frequently in the corpora. The frequency ranking of these verbs within Portuguese according to a frequency dictionary (Davies 2008) is given in Table 26. I also include the figures for separate entries in the frequency dictionary for forms identical to the irregular past participle, which could have affected the frequencies listed for the verbs.

Verb	**Frequency ranking**	**Separate related entries**
aceitar	368	
acender	3397	*aceso*: 4276
gastar	1270	*gasto*: 1911
limpar	1421	*limpo*: 3777
matar	431	*morto*: 575
pagar	274	
eleger	1005	
entregar	690	
prender	1437	*preso*: 483
suspender	2320	*suspenso*: 1927

Table 26 Frequencies of selected Portuguese verbs

In the following sections, data from two corpora are presented: in section 5.1.4.1, the 45-million word *Corpus do Português: Genre/Historical* (Davies and Ferreira 2006), and in section 5.1.4.2, the

Dicionário de Verbos do Português Medieval, which lists all the verb forms attested in the *Corpus Informatizado do Português Medieval* (Xavier, Crispim, and Vicente n.d.). The diachronic corpus data reveal complexities in the historical path of the two past participle forms which would not otherwise be obvious.

5.1.4.1 Corpus do Português: Genre/Historical

The corpus was searched for all instances of the auxiliaries *ter* 'have' and *ser* 'be' with both the regular and irregular past participles of the selected ten verbs. The results are presented in tables and graphs below, allowing us to identify the relative frequencies of forms over time. In tables, the irregular form is always given on the left, and the regular form on the right. Grey cells indicate that no forms were found in that century of the corpus with that particular verb and auxiliary verb combination. In each case, the more frequent form (whether regular or irregular) is marked in bold. In graphs, the irregular form is presented in black at the base of each bar, and the regular form in grey from the top of each bar.

The results for the verb *matar* should be taken with caution, because they are likely to include past participles of the identical form *morto* from the verb *morrer* 'die': in the context with *ter* 'have' it is impossible to distinguish 'have died' from 'have killed' other than by manual disambiguation, which is impractical given the number of tokens for this verb. Caution is also needed when interpreting the results for any verb with very few tokens (such as *limpar*).

	Aceitar					Entregar				
TER	aceite/o	%	aceitado	%	Total	entregue	%	entregado	%	Total
1200s	0		0		0	0		0		0
1300s	0		0		0	0		0		0
1400s	0		0		0	0		0		0
1500s	0	*0*	**2**	***100***	2	**5**	***83***	1	*17*	6
1600s	0	*0*	**6**	***100***	6	**3**	***75***	1	*25*	4
1700s	0	*0*	**3**	***100***	3	**3**	***75***	1	*25*	4
1800s	1	*5*	**20**	***95***	21	8	*36*	**14**	***64***	22
1900s	**50**	***60***	33	*40*	83	**23**	***77***	7	*23*	30
	Gastar					Limpar				
TER	gasto	%	gastado	%	Total	limpo	%	limpado	%	Total
1200s	0		0		0	0		0		0
1300s	0		0		0	0		0		0
1400s	0	*0*	**2**	***100***	2	**1**	***100***	0	*0*	1
1500s	1	*6*	**16**	***94***	17	**2**	***100***	0	*0*	2
1600s	0	*0*	**16**	***100***	16	0		0		0
1700s	3	*38*	**5**	***63***	8	0		0		0
1800s	**32**	***91***	3	*9*	35	0	*0*	**1**	***100***	1
1900s	**35**	***92***	3	*8*	38	**4**	***80***	1	*20*	5
	Matar					Pagar				
TER	morto	%	matado	%	Total	pago	%	pagado	%	Total
1200s	0		0		0	0		0		0
1300s	**2**	***100***	0	*0*	2	0	*0*	**1**	***100***	1
1400s	**4**	***80***	1	*20*	5	**1**	***100***	0	*0*	1
1500s	**22**	***100***	0	*0*	22	**16**	***84***	3	*16*	19
1600s	**13**	***100***	0	*0*	13	**15**	***94***	1	*6*	16
1700s	**18**	***95***	1	*5*	19	**4**	***100***	0	*0*	4
1800s	18	*44*	**23**	***56***	41	**30**	***91***	3	*9*	33
1900s	27	*33*	**56**	***67***	83	**63**	***97***	2	*3*	65

Table 27 Frequencies of past participle forms of verbs in the first conjugation with *ter* 'have' (*Corpus do Português* Genre/Historical)

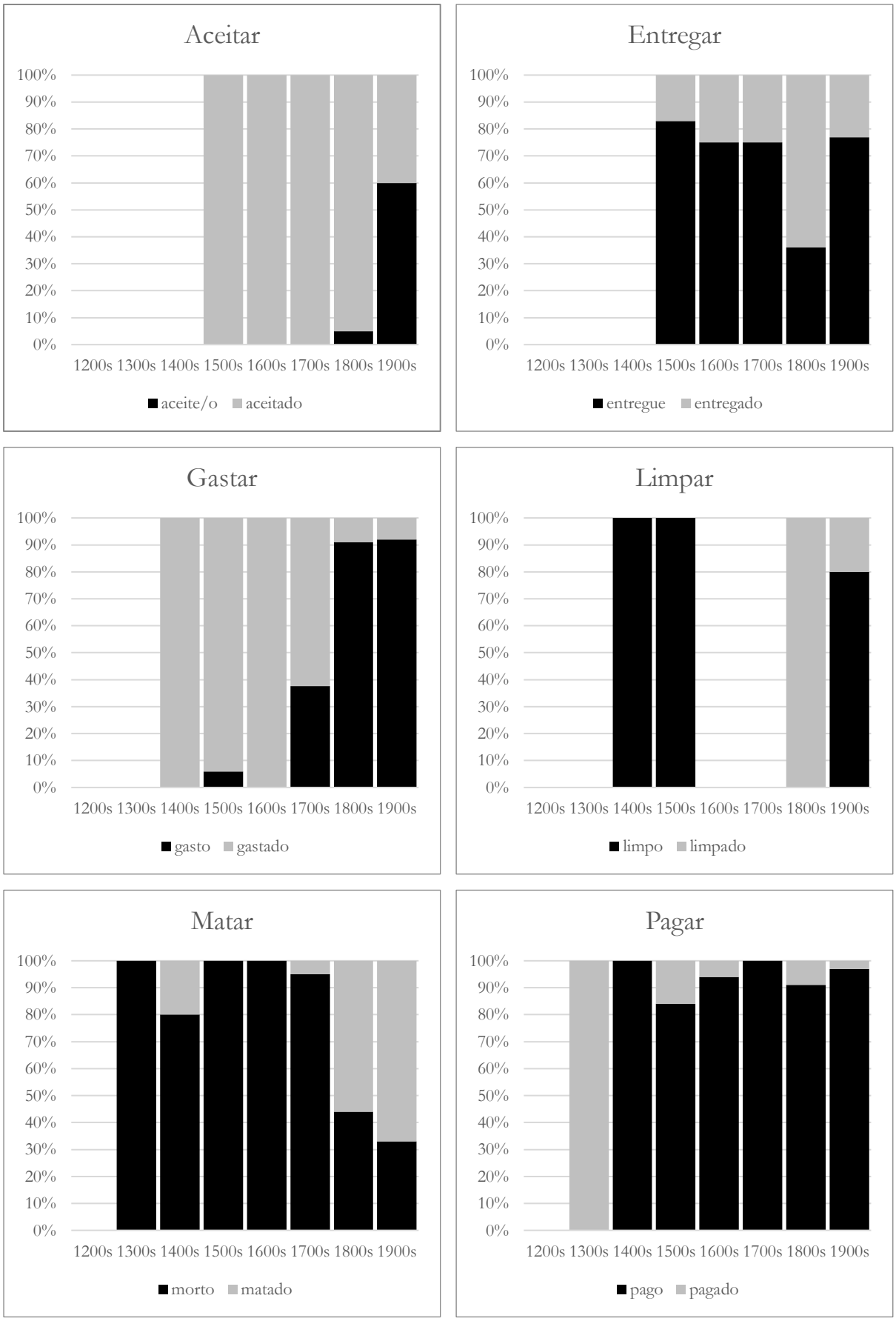


Figure 9 Relative frequencies of irregular and regular past participle forms for first conjugation verbs with auxiliary *ter* 'have'

We see different patterns in the data, but the general tendency of first conjugation verbs in the active context with auxiliary *ter* 'have' (Table 27, Figure 9) is towards the loss of the regular form, in favour of the irregular form. This change has taken place at different times for different verbs. For the verbs *aceitar, gastar* and *pagar*, the irregular now dominates, a change shown particularly clearly for *aceitar* and *gastar*. Results for *limpar* and *matar* are inconclusive, while those for *entregar* suggest a more complex situation in which the short form dominates in every century except the 1800s. This could indicate a short-lived preference for the long form in the norm of that century, but further research which explores beyond this sample is needed to draw final conclusions.[86]

[86] It is possible that this is an artefact of the texts selected within the corpus. For example, the data for the 1800s could be skewed towards an author who unexpectedly favoured the form *entregado*.

	Acender					Eleger				
TER	aceso	%	acendido	%	Total	eleito	%	elegido	%	Total
1200s	0		0		0	0		0		0
1300s	0		0		0	0		0		0
1400s	0		0		0	0		0		0
1500s	0		0		0	1	*33*	**2**	***67***	3
1600s	0		0		0	0		0		0
1700s	0		0		0	0	*0*	**1**	***100***	1
1800s	0	*0*	**13**	***100***	13	0	*0*	**1**	***100***	1
1900s	2	*15*	**11**	***85***	13	4	*50*	4	*50*	8
	Prender					Suspender				
TER	preso	%	prendido		Total	suspenso	%	suspendido	%	Total
1200s	**2**	***100***	0	*0*	2	0		0		0
1300s	**20**	***100***	0	*0*	20	0		0		0
1400s	**14**	***100***	0	*0*	14	0		0		0
1500s	**20**	***100***	0	*0*	20	0		0		0
1600s	**4**	***100***	0	*0*	4	1	*33*	**2**	***67***	3
1700s	**1**	***100***	0	*0*	1	0		0		0
1800s	**10**	***56***	8	*44*	18	1	*33*	**2**	***67***	3
1900s	4	*50*	4	*50*	8	1	*14*	**6**	***86***	7

Table 28 Frequencies of past participle forms of verbs in the second conjugation with *ter* 'have'

(*Corpus do Português* Genre/Historical)

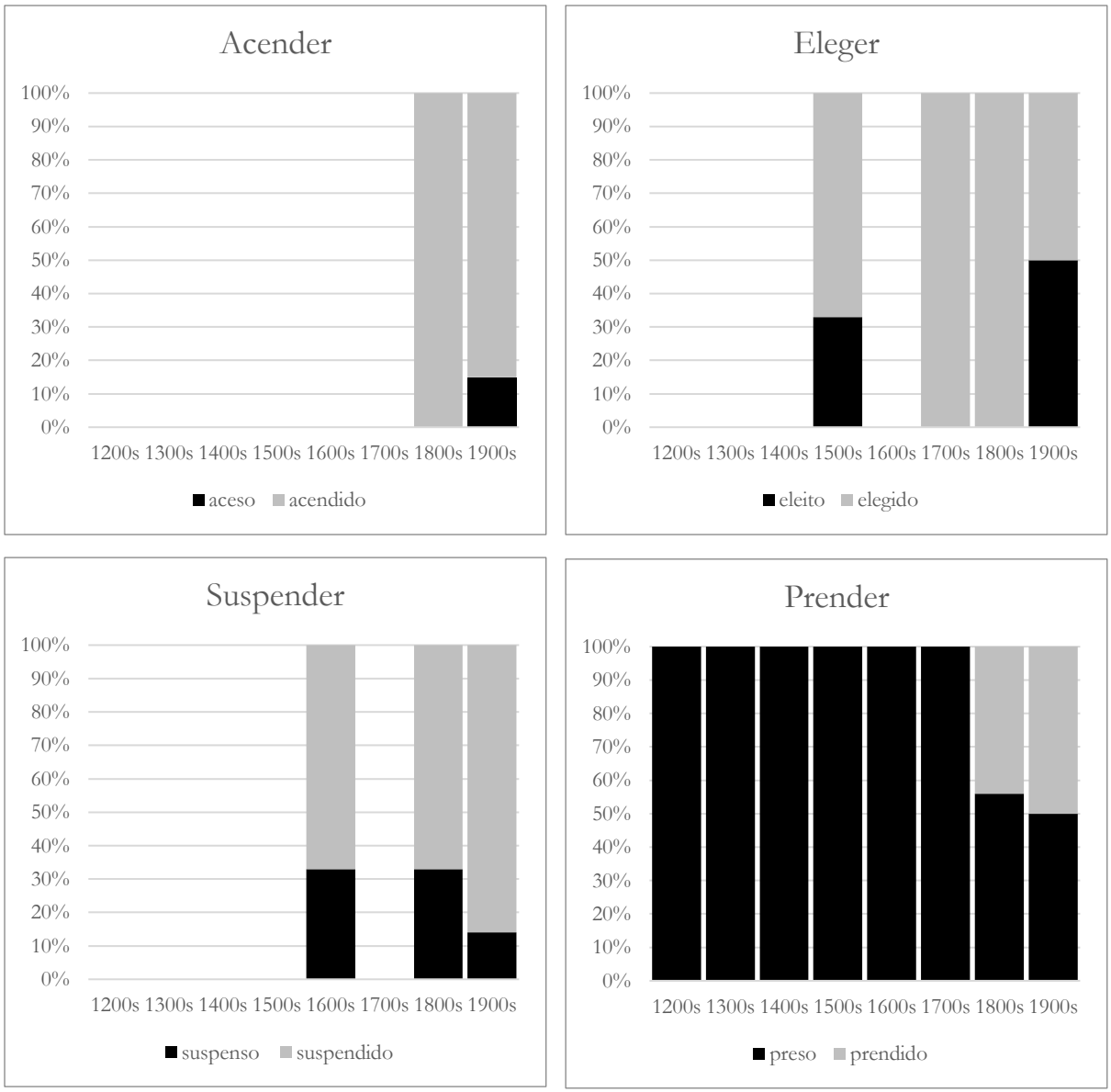


Figure 10 Relative frequencies of irregular and regular past participle forms for second conjugation verbs with auxiliary *ter* 'have'

The results for second conjugation verbs in the active context (auxiliary *ter* 'have': Table 28, Figure 10) are also complex. The verbs *acender* and *suspender* show a preference for the regular form with *ter* (when attested). The verbs *eleger* and *prender* show convergence from opposite positions: a preference for the regular form of *eleger* and for the irregular form of *prender* has been erased, with the 1900s showing 50% usage of both forms. The data suggest that the form *prendido* is a recent innovation, whereas *elegido* has a much longer history. In general, there seems to be more variability in this group of verbs than in the first conjugation, in this context. This could be due to the tension

between the expected forms for second conjugation verbs (irregular forms, as inherited) and the expected forms for the active context (regular forms).

Next I consider the passive context with *ser* 'be', starting with the first conjugation.

	Aceitar					Entregar				
SER	aceite/o	%	aceitado	%	Total	entregue	%	entregado	%	Total
1200s	0		0		0	1	*10*	**9**	***90***	10
1300s	0		0		0	11	*41*	**16**	***59***	27
1400s	0		0		0	**30**	***97***	1	*3*	31
1500s	**7**	***88***	1	*13*	8	**26**	***100***	0	*0*	26
1600s	**6**	***55***	5	*45*	11	**13**	***100***	0	*0*	13
1700s	**3**	***75***	1	*25*	4	**13**	***100***	0	*0*	13
1800s	**25**	***100***	0	*0*	25	**67**	***100***	0	*0*	67
1900s	**155**	***100***	0	*0*	155	**305**	***100***	0	*0*	305
	Gastar					Limpar				
SER	gasto	%	gastado	%	Total	limpo	%	limpado	%	Total
1200s	0		0		0	0		0		0
1300s	0		0		0	**1**	***100***	0	*0*	1
1400s	1	*50*	1	*50*	2	**16**	***100***	0	*0*	16
1500s	2	*29*	**5**	***71***	7	**18**	***100***	0	*0*	18
1600s	**2**	***100***	0	*0*	2	**4**	***100***	0	*0*	4
1700s	0		0		0	**5**	***100***	0	*0*	5
1800s	**4**	***100***	0	*0*	4	**7**	***100***	0	*0*	7
1900s	**28**	***100***	0	*0*	28	**13**	***93***	1	*7*	14
	Matar					Pagar				
SER	morto	%	matado	%	Total	pago	%	pagado	%	Total
1200s	**36**	***100***	0	*0*	36	0	*0*	**4**	***100***	4
1300s	**210**	***100***	0	*0*	210	2	*11*	**16**	***89***	18
1400s	**266**	***100***	1	*0*	267	**15**	***65***	8	*35*	23
1500s	**171**	***98***	4	*2*	175	**7**	***88***	1	*13*	8
1600s	**48**	***100***	0	*0*	48	**4**	***100***	0	*0*	4
1700s	**28**	***100***	0	*0*	28	**6**	***100***	0	*0*	6
1800s	**76**	***100***	0	*0*	76	**27**	***100***	0	*0*	27
1900s	**280**	***100***	1	*0*	281	**146**	***100***	0	*0*	146

Table 29 Frequencies of past participle forms of verbs in the first conjugation with *ser* 'be' (*Corpus do Português* Genre/Historical)

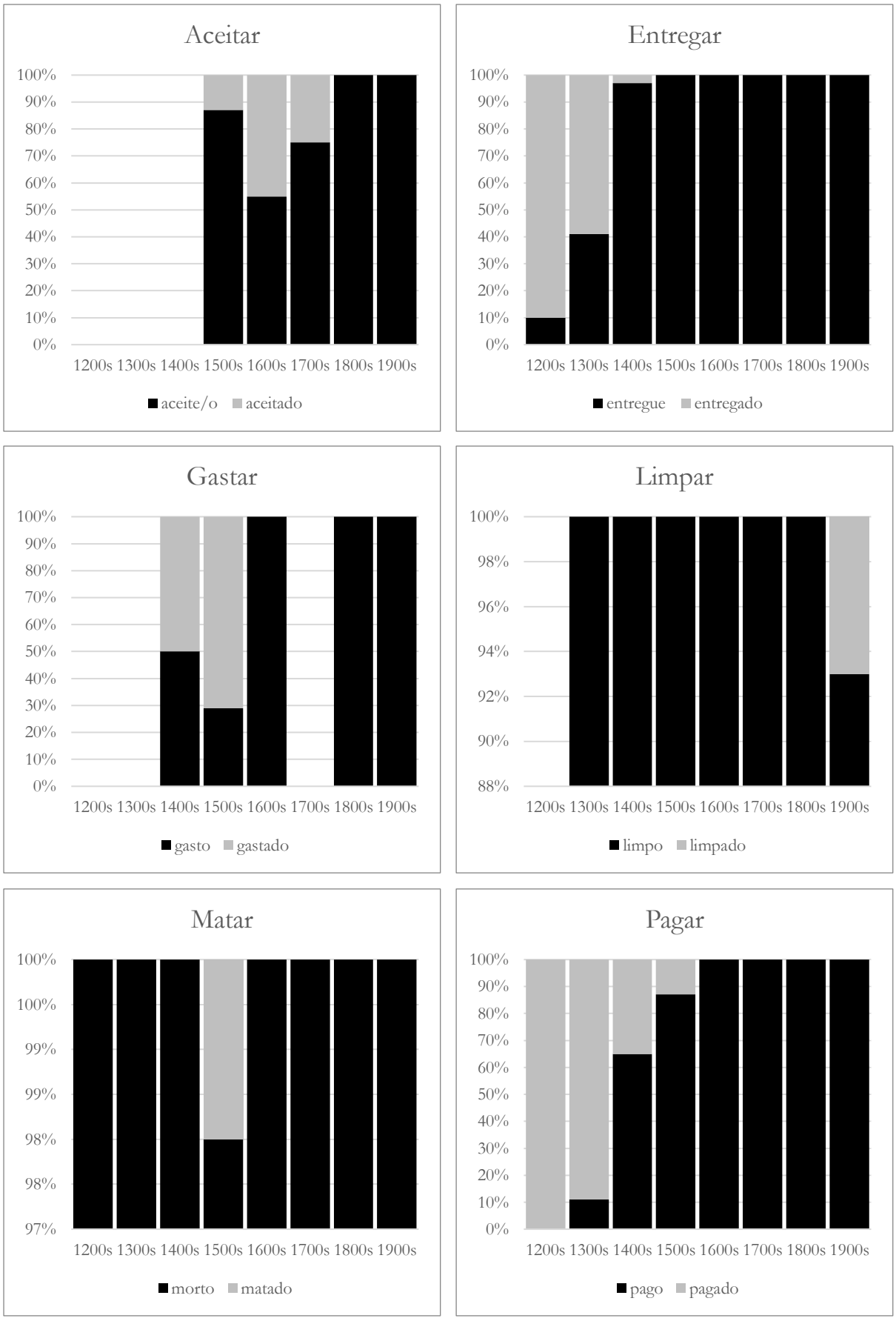


Figure 11 Relative frequencies of irregular and regular past participle forms for first conjugation verbs with auxiliary *ser* 'be'

The results for the passive context (with auxiliary *ser*) are more straightforward. Some first conjugation verbs (Table 29, Figure 11) including *gastar, pagar* and *entregar* preferentially show the regular form with *ser* at some point in their history, but the irregular form is now preferred, which is compatible with the suggestion made above that the irregular form of these verbs was innovated within the history of Portuguese. The move towards the irregular form is attested around the 1400s to 1600s for these verbs, although with some variation.

This pattern is not seen in *aceitar* which inherited both regular and irregular forms. For *limpar* 'clean' and *matar* 'kill' the situation is made more complex by the fact that their irregular form is identical to uncontroversial adjectives *limpo* 'clean' and *morto* 'dead' which are likely to be found in the copular context.

Second conjugation verbs (Table 30, Figure 12), namely *prender, suspender, acender* (and arguably *eleger* given the small number of tokens from the 1300s and 1400s) have always favoured their (inherited) irregular past participle form in this context.

	Acender					Eleger				
SER	aceso	%	acendido	%	Total	eleito	%	elegido	%	Total
1200s	0		0		0	0		0		0
1300s	0		0		0	0	*0*	**1**	***100***	1
1400s	**2**	***67***	1	*33*	3	0	*0*	**2**	***100***	2
1500s	**1**	***100***	0	*0*	1	**9**	***82***	2	*18*	11
1600s	0		0		0	**21**	***84***	4	*16*	25
1700s	0		0		0	**4**	***100***	0	*0*	4
1800s	0		0		0	**34**	***100***	0	*0*	34
1900s	**1**	***100***	0	*0*	1	**799**	***100***	0	*0*	799
	Prender					Suspender				
SER	preso	%	prendido	%	Total	suspenso	%	suspendido	%	Total
1200s	**3**	***100***	0	*0*	3	0		0		0
1300s	**54**	***100***	0	*0*	54	0		0		0
1400s	**69**	***100***	0	*0*	69	**3**	***100***	0	*0*	3
1500s	**54**	***100***	0	*0*	54	**3**	***100***	0	*0*	3
1600s	**7**	***100***	0	*0*	7	**1**	***100***	0	*0*	1
1700s	**5**	***100***	0	*0*	5	0		0		0
1800s	**106**	***100***	0	*0*	106	**1**	***100***	0	*0*	1
1900s	**505**	***100***	0	*0*	505	**65**	***100***	0	*0*	65

Table 30 Frequencies of past participle forms of verbs in the second conjugation with *ser* 'be' (*Corpus do Português* Genre/Historical)

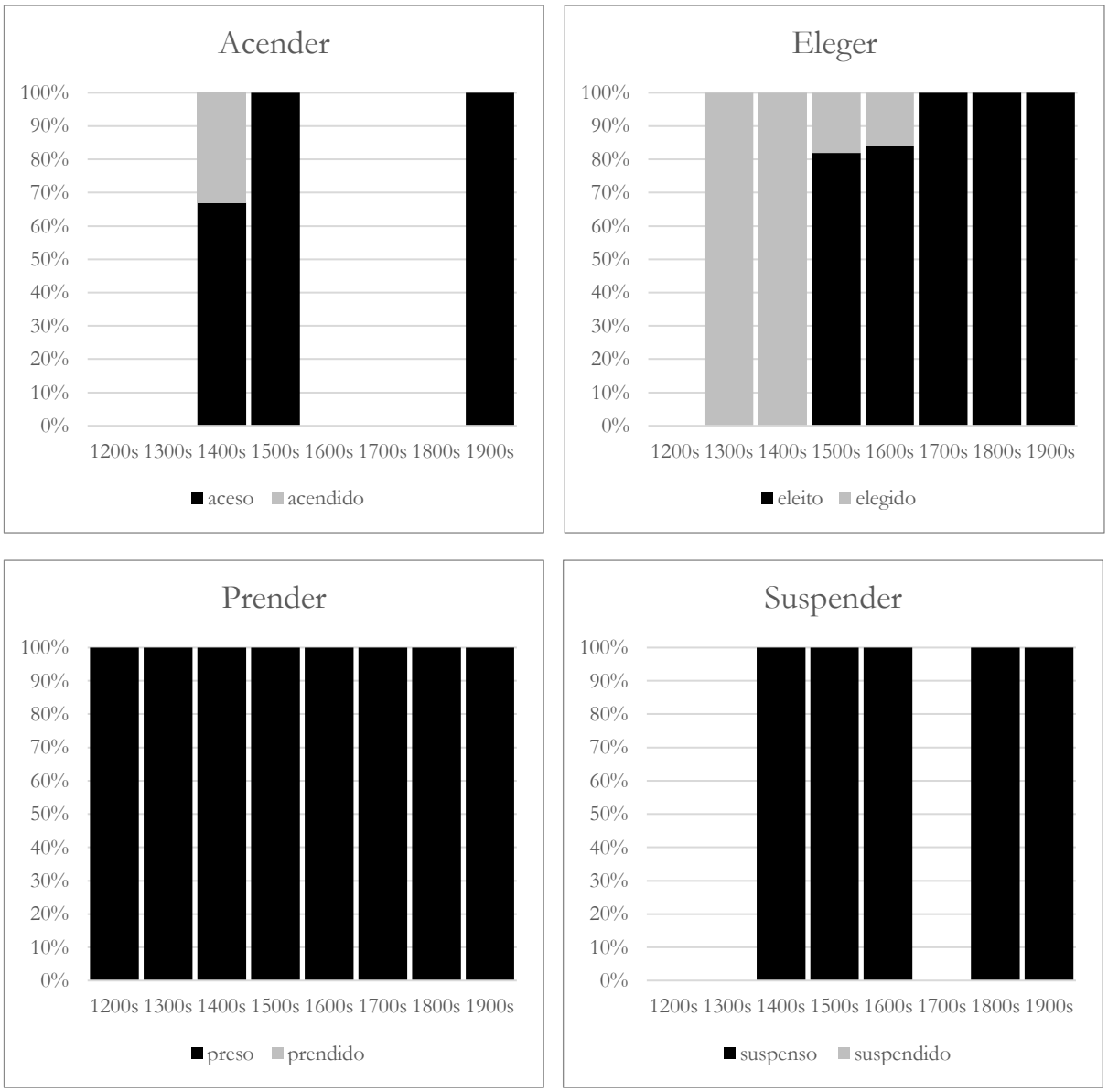


Figure 12 Relative frequencies of irregular and regular past participle forms for second conjugation verbs with auxiliary *ser* 'be'

As with much diachronic work, there are serious limitations as to the number of attestations and the quality of the data represented in these tables and figures. Most importantly, there is no guarantee that forms attested with *ser* 'be' are really acting as past participles. It is not straightforward to disambiguate the functions of the forms, which can be fully adjectival.

With this in mind, some conclusions can be drawn. Most strikingly, the data suggest that there were never well-defined uses of the regular and irregular forms across every verb. That is to say, it is possible that for some verbs, there was always an element of variability in which past participle was chosen with *ser*, as well as with *ter*. Indeed, if we bear in mind the fact that these texts represent

many authors from different regions of Portugal, produced in distinct contexts over many centuries, such variation is perhaps to be expected.

Without a detailed qualitative analysis of the texts that make up this historical corpus, it is unclear whether variability within verbs and within centuries is due to variation in usage between individual authors, or due to genuine divergence in the grammatical rules in operation. A comprehensive investigation of this would provide greater clarity about the true diachronic situation of these forms.

As can be seen from the data so far, there is no overall schema for the diachronic development of past participle forms. We are instead dealing with considerable variation which can be traced back to Latin and reflects 'verb-by-verb' processes of analogical change. This is also borne out by the data from the *Dicionário de Verbos do Português Medieval*, as discussed in the next section.

5.1.4.2 Dicionário de Verbos do Português Medieval

The *Dicionário de Verbos do Português Medieval* takes the form of a dictionary of verbs, with a list of every wordform used in the *Corpus Informatizado do Português Medieval*. The same ten verbs are studied here as in the previous section. It was not possible to obtain examples of each form in context, so the list below (Table 31) indicates that a given form existed during the medieval period, along with the century or centuries in which it is attested. The dictionary lists forms of each number/gender agreement (MSG/MPL/FSG/FPL) but for simplicity, only MSG forms are reported here.

As can easily be seen, not all forms are attested in the corpus. Some verbs are entirely absent (*suspender*, *pagar*), and for other verbs, the short form may be missing (*alimpar, matar*) or the long form (*prender*).

Verb	**Regular form**	**Irregular form**
aceitar	aceptado (15th)	acceito (15th), acçepto (16th) açepto (16th)
acender	açendudo (14th), acendudo (14th), accendido (15th)	aceso (13/14th), açeso (15th), acesso (15th)
gastar	gastado (15th), guastado (15th)	gasto (15th)
limpar (alimpar)	alinpado (14th), alimpado (15th)	-
matar	matado (?)	-
pagar	-	-
entregar	entregado (13th), ent(re)guado (13th), entergado (15th)	entrege (14th), emtregue (15th)
eleger	emlegido (14th), enlegido (14th), elegido (14th), enligyudo (14th), ẽlegido (15th)	enleito (14th), eleito (14th), emleito (15th), ẽleyto (15th), electo (15th)
prender	-	preso (13th/14th) presso (14th) p(re)so (14th) p(re)sso (14th)
suspender	-	-

Table 31 Past participle forms included in the *Dicionário do Português Medieval*

These data generally reflect the findings from the *Corpus do Português*, in that there are no striking contradictions in the attestations of either regular or irregular forms between the two corpora, other than the expected gaps in attestations.

Taking the different sources of data together, we now have a reasonable insight into the different trajectories of a selection of Portuguese verbs in the history of European Portuguese. In the next section, I turn towards Brazilian Portuguese.

5.1.5 Divergence across varieties: double past participles in Brazilian Portuguese

In Chapter 4 we saw briefly that there are differences between regional varieties of Portuguese in the double past participle phenomenon, although the details and extent of these differences are poorly understood. This synchronic regional variation reflects divergent diachronic development, making it necessary to consider in this chapter some of the different possibilities of development that past participle forms could undergo across different varieties, leading to modern variation. In this section I set out some possible differences between Brazilian and European Portuguese,[87] preparing the ground for a later explanation in section 5.3 of how these differences can be represented in the LFG-PFM model.

To recapitulate the relevant details from Chapter 4, there has been little work specifically on the differences between European and Brazilian Portuguese in the double past participle phenomenon. Schwenter et al. (2019) present a small amount of data from the *Corpus do Português* suggesting that irregular past participles are used more in active contexts in European Portuguese than in Brazilian Portuguese. The data from Schwenter et. al (2019: 75, Table 2) are summarised in Table 32.

[87] I do not claim to present an authoritative account of the extent and details of the differences between the macro-varieties of Brazilian and European Portuguese, nor do I consider in detail the possible regional or social variation within the macro-varieties. Further detailed empirical research into the distribution and use of different past participle forms across both Brazilian and European Portuguese would be needed to provide such analyses.

	Brazil	**Portugal**
Irregular Participle	164 (53%)	274 (68%)
Regular Participle	145 (47%)	128 (32%)
Total	*309*	*402*

Table 32 Frequencies of participle form across dialects in 20th century data from *Corpus do Português*

The usefulness of these data can be questioned, given that they could simply reflect differences in the frequencies of particular verbs within the corpus selections: even if Brazilian Portuguese and European Portuguese verbs behaved identically in active contexts (with some verbs being disposed towards the irregular participle and others towards the regular participle), different overall frequencies of the irregular versus regular forms would be possible if some verbs occur more frequently in the Brazilian Portuguese data than the European Portuguese.

One possible interpretation of the data is therefore that exactly the same verbs have regular and irregular past participle forms in both varieties. Another possible interpretation is that some (or all) verbs are more likely to have a regular past participle form in Brazilian Portuguese, but an irregular past participle form in European Portuguese. Since Schwenter et al. (2019) do not provide a breakdown of the data verb-by-verb separately for Brazilian Portuguese and European Portuguese, the correct interpretation is unknown.

To take a further step back, it is difficult to extrapolate anything about the double past participle phenomenon (in the sense of the term used throughout this thesis) from these data. Schwenter et al.'s (2019) perspective is that verbs with variation between regular and irregular past participles in the active context are exemplars of the double past participle phenomenon, a claim that differs from my understanding of the phenomenon, which I believe better reflects previous work in this area. In my description, a 'true' double past participle verb would have *only* the regular form in the active context, and *only* the irregular form in the passive context. Such a verb would not be counted as a double past participle verb in Schwenter et al.'s methodology, because only one form would

be used in the active context. This further reduces the usefulness of these results for a principled account of the differences in the double past participle phenomenon between European and Brazilian Portuguese.

Setting aside these difficulties, I now consider a feature of Brazilian Portuguese that shows clear divergence from European Portuguese: the innovative past participle forms.

The existence in some Brazilian Portuguese dialects of innovative double past participle forms has been the subject of some attention, as summarised above in section 4.3.3. These forms represent the extension of athematic past participle forms, identical to the first person singular present indicative verb form, primarily in the first conjugation, but also in other conjugations. Some examples from Chagas de Souza (2007, 126) were given in Chapter 4 and are repeated here for convenience and renumbered [my gloss, original translation].

Innovative (non-standard) past participles in the first conjugation:

(151) *chegar* 'arrive' > *chego* 'arrived'

A carta tinha chegado/chego

the letter had arrived.REGPTCP /IRREGPTCP

'The letter had arrived'

Innovative (non-standard) past participles in other conjugations:

(152) *trazer* 'bring' > *trago* 'brought'

O professor tinha trazido/trago o livro

the teacher had brought.REGPTCP /IRREGPTCP the book

'The teacher had brought the book'

(153) *pedir* 'ask' > *peço* 'asked'

O presidente tinha pedido/peço paciência

the president had asked.for.REGPTCP /IRREGPTCP patience

'The president had asked for patience'

There is little information available on the exact use of these irregular forms, in particular their distribution between active and passive contexts: one key paper on the subject notes that 'there are no researches mapping the possible uses of innovative participles in BP' (Scher, Lunguinho, and Rodero-Takahira 2014, 47), before demonstrating that the participles are used in both active perfect and passive contexts.[88] The same authors previously argued (Scher, Lunguinho, and Rodero-Takahira 2013) that the innovative participles were only used in active contexts, before re-evaluating this claim and stating that 'one cannot ignore that the alleged asymmetries in the distribution of innovative participles regarding active and passive contexts might not be real' (Scher, Lunguinho, and Rodero-Takahira 2014, 57). One possible analysis of the data is that Brazilian Portuguese innovative past participles have exactly the same properties as regular past participles, and that they can wholly replace regular past participles across active and passive contexts for some speakers. Alternatively, the innovative past participles may be first associated with the active context originally, and then extended to passive contexts.

Although it is argued that there are some Brazilian Portuguese 'dialects' in which the innovative forms are used (Scher, Lunguinho, and Rodero-Takahira 2013; 2014), what exactly these 'dialects' are is not considered. They could be small socially or regionally limited varieties which are stable

[88] The paper argues that the use of innovative participles in passive contexts is more recent than the use of innovative participles in active contexts, but no substantiation is given for this claim other than the authors' perceptions.

and distinctive, or it is possible that the evidence of the existence of these innovative past participles, although apparently somewhat widespread, reflects more idiolectal variation. This could be reflective of change in progress, as some verbs lose their regular past participle forms in favour of irregular forms. In either case, this would represent a difference between European Portuguese and Brazilian Portuguese.

A useful study of innovative past participles in Brazilian Portuguese is presented by Teixeira da Silva (2008), focusing on Portuguese in the South of Brazil. Teixeira da Silva presents data from the VARSUL corpus, including 24 speakers from the cities of Porto Alegre, Florianópolis and Curitiba, as well as a behavioural experiment with 22 native speakers from the South of Brazil with university-level education. Speakers made a forced choice between either the regular or the irregular past participle form for a mixture of real and invented Portuguese verbs. Since verbs were tested either in the passive context or in the active context, it is unclear whether there is an active/passive distinction: however, it is notable that there was considerable disagreement between speakers on the preferred form of many real Portuguese verbs, as well as the invented verbs (Teixeira da Silva 2008, 84). For example, the verb *cobrir* 'cover' has only an irregular form *coberto*, but a regularised form **cobrido* was also presented and chosen by two participants, which Teixeira da Silva states was unexpected (2008, 54). This suggests that either there is far more variation than previously realised, or that the forced choice methodology contributed to misunderstandings of the task or uncooperative behaviour of some participants. More generally, there was an unexpectedly high use of irregular past participle forms in a test designed to study productivity of the different forms. This is attributed to 'hypercorrection' as irregular past participles are perceived as higher status by some speakers.

The corpus data are more interesting for our purposes. Teixeira da Silva provides an appendix containing all the relevant forms extracted from the corpus with both the auxiliary verb and the past participle (2008, 86–93). Interestingly, the vast majority of the verbs are attested either in the

active context or in the passive context, but not both. There is some variation in the forms that are attested: both *chego* and *chegado* are attested in the active context for *chegar* 'arrive', with the former being an innovative Brazilian Portuguese form. Both *ganho* and *ganhado* are attested in the active context for *ganhar* 'earn'. For *morrer* 'die', *morrido* is attested with *ter* 'have' and *morto* with auxiliary *estar* 'be': however, it is unlikely that *morto* is genuinely a past participle in this context (see section 5.3.4 for further discussion). Otherwise only one past participle form is attested per verb, including the standard irregular forms *entregue* for *entregar* 'hand over' and *eleito* for *eleger* 'choose', and the innovative Brazilian form *pego* for *pegar* 'take, grab'.[89]

In sum, the VARSUL data provides no evidence in favour of a double past participle phenomenon where the use of regular/irregular past participles is conditioned by active/passive contexts, in the educated variety of Southern Brazilian Portuguese represented in the corpus. We cannot rule out that some Brazilian Portuguese varieties do have such a distinction, and indeed it may be the cause for some Southern Brazilian Portuguese speakers, but the VARSUL data does not show it. On the other hand, the data are certainly compatible with an analysis in which there is no such pattern. This line of analysis is taken up in section 5.3.3.2.

To sum up this survey, we see two possible differences between double past participles in European Portuguese and in Brazilian Portuguese. There is good evidence for the existence of innovative irregular past participle forms in some speakers' Brazilian Portuguese which are not found in European Portuguese: past participles of the truncated pattern are extended analogically

[89] Teixeira da Silva also lists *está cheio* 'it is full' under the verb *encher* 'fill' and *era bem quisto* 'he/she was well-liked' under *querer* 'want', but both *cheio* and *benquisto* (standard orthography) are straightforward adjectives in modern Portuguese, suggesting the criteria used to evaluate the forms may have lacked precision.

to a wider set of verbs in Brazilian Portuguese than in European Portuguese. These innovative past participle forms are notable as they provide evidence for past participles being formed by analogy with first person singular present indicative forms. The second difference relates to the distribution of regular and irregular past participles across active and passive contexts. On this matter, the evidence is poor. However, there is some evidence that the alignment between active contexts and regular past participle and passive contexts and irregular past participles made by at least some speakers for some verbs in standard European Portuguese is not operative in at least some Brazilian Portuguese varieties.

As described in section 4.3.2, Bonami and Luís present examples from the CETEMPúblico corpus for one verb *encarregar* 'entrust' suggesting that the active/passive split for regular/irregular past participles may no longer exist in European Portuguese, with the two forms being in free variation. Although this may be the case for some speakers and some verbs, there is good evidence that the phenomenon does still exist to a degree in European Portuguese. I therefore generally treat the loss of the active/passive distinction as a Brazilian Portuguese phenomenon, although future detailed empirical work would clarify the extent to which it is also found in European Portuguese. For example, this would clarify whether any individual speakers have both forms of the verb *envarregar* (or other verbs) in 'free variation', or whether there is simply variation across speakers as to whether they have just the regular or irregular form (in both active and passive contexts) or maintain both forms distinguished by context. Within the PFM system, it is straightforward to account for the loss of either the regular or irregular form of one verb (and thus the removal of that verb from the class of verbs that demonstrate the double past participle phenomenon), as I show in section 5.3.2 below. Equally, I also provide an account of a more general loss of the active/passive distinction, in section 5.3.3.2.

Turning back to the two major differences between European and Brazilian Portuguese past participles as characterised here (the innovative past participles, and the lack of active/passive

distinction), accounting for the Brazilian Portuguese forms requires alterations to the PFM rules given for European Portuguese in Chapter 4, as is expected. The PFM rules are only intended to capture one variety, but can easily be adjusted to capture morphological variation across linguistic systems. In this respect, purely morphological variation and change in morphological forms can be captured in the model without syntactic consequences in the LFG analysis, which is desirable for phenomena in which the morphology changes independently of the syntax. On the other hand, instances when the morphology and syntax are closely linked can also be captured in the same framework.

Comparing across the PFM analyses that we develop for each variety gives us new insights into the kinds of divergent development that actually occur, such as the gain or loss of certain types of rules, or particular changes in inflection class membership or theme vowels. This permits a deeper understanding of the possibilities of morphological change and divergence between varieties. Section 5.3.3 illustrates how the PFM rules can account for the differences seen in past participles across Brazilian and European Portuguese.

5.1.6 Summary of diachronic trajectories

As we have seen throughout this section, there is substantial variation in the diachronic trajectories of the past participles of different verbs, dating back to Latin and continuing to modern patterns of innovation. Figure 13 below summarises the general trajectories that are possible for Portuguese verbs in terms of the forms they inherited from Latin and those they have in modern Portuguese. The representations are simplified for clarity: verbs which innovated one form and lost another may have gone through an intermediate stage as double past participle verbs, which is not represented in these figures.

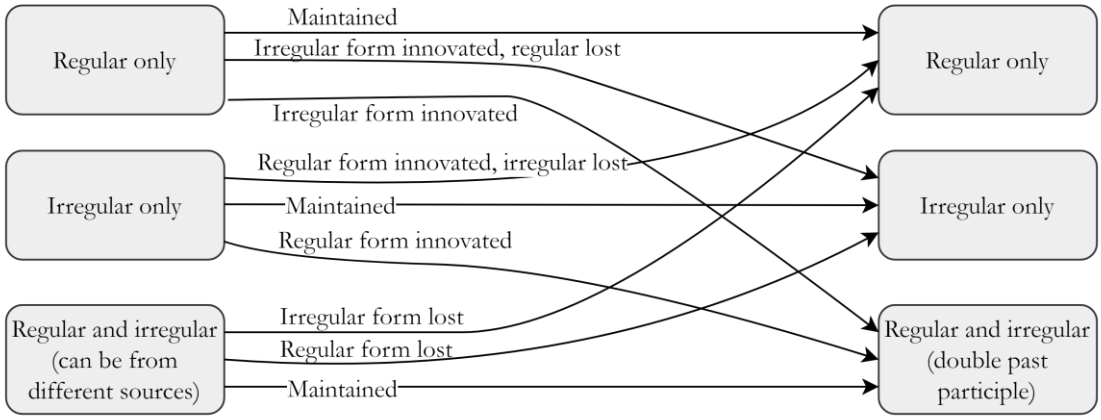


Figure 13 Possible trajectories of Portuguese verbs' past participle forms

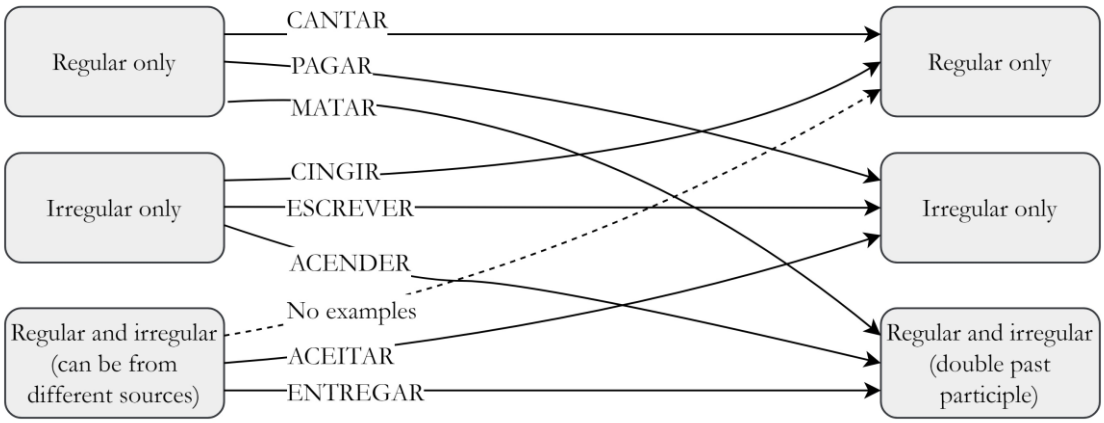


Figure 14 Portuguese verbs exemplifying possible trajectories of the past participle forms[90]

Various factors are involved in the variation seen across verbs. One particularly striking difference is the distinction in Latin between first conjugation verbs, which always had thematic past participles, and verbs in other conjugations, some of which had thematic and some of which had

[90] Although I am not aware of examples of verbs which used to have a regular and an irregular past participle form and now only have a regular form, there is no a priori expectation that this would be the case, and indeed examples may exist. Irregular forms can certainly fall out of use in favour of regular forms, as we saw for *absolver* 'absolve', *cingir* 'gird' and *defender* 'defend' in section 5.1.3.

athematic past participles. The consequences of this split were far-reaching and can still be seen in the morphological patterns of modern Portuguese verbs.

Verbs in the first inflection class generally inherited a regular past participle from Latin. Some verbs in the class subsumed an existing adjectival form as an irregular past participle, e.g. *gasto* for *gastar* 'spend' in place of *gastado*. Other first conjugation verbs innovated an irregular past participle on the same pattern, consisting of the regular verb stem with no theme vowel and no consonant /d/ such as *pago* for *pagar* 'pay', where the attested Latin form was *pacatus*. In some of the verbs that used to have two past participle forms, the irregular form has fully replaced the inherited regular form. The extension of this type of irregular form can be seen in modern Brazilian innovative past participles too, as discussed in section 5.1.5.

Portuguese verbs in the second and third conjugations were more likely to inherit only an irregular (athematic) past participle form from Latin. Some of these verbs have subsequently innovated regular past participles by analogy with the existing thematic past participles of other verbs in these inflection classes. As the diachronic corpus data make clear, though, there is no single pathway of development that explains the histories of all these verbs or their regular and irregular forms. Accounting for the contexts in which the different forms were used (active versus passive) is particularly difficult. Instead, we must recognise that we have only a partial view of the past, and the model I propose here can only represent the visible data.

Before I present PFM rules that account for some of the diachronic changes seen in the past participles (section 5.3), in the next section I consider the types of adaptations that can be made within the PFM system to capture diachronic variation, considering the role of major inflection classes, lexical specification and analogy.

5.2 Analysing diachronic change

This section considers the evidence presented in the previous section about the development of different forms (regular and irregular) in the history of Portuguese, and works through some of the questions involved in modelling these historical developments. This gives an improved foundation for the presentation in section 5.3 of a PFM account of historical changes to past participles, represented through static analyses capturing minimal differences in Portuguese verbal morphology at different points in time.

As we saw in the previous section, inflection class plays a role in determining the types of changes that a verb might undergo. However, I discuss in section 5.2.1 some reasons why integrating the traditional inflection classes (first, second and third) as factors in diachronic change into our model and theoretical rules might be undesirable.

In section 5.2.2, I discuss the role of lexical specification in a PFM model of change, focusing on the different ways that lexical information is represented in PFM, and discussing the implications of this on the ways in which diachronic developments are modelled. Finally, in section 5.2.3 I briefly consider the role and representation of analogical processes in this model of morphological change.

5.2.1 The role of major inflection classes in the diachrony of the past participles

The data presented in section 5.1 show clearly that there are different tendencies of development across the major inflection classes of Portuguese, that is the first, second and third conjugations represented in my PFM account as **C1**, **C2** and **C3**. In particular, there are tendencies in the patterns by which irregular (athematic) or regular (thematic) past participle forms are inherited from Latin versus innovated within Portuguese. However, the data also make clear that not all verbs in a single inflection class behave the same way.

A formal model of the diachronic data should not overgeneralise the patterns that we see across inflection classes. Only a small number of Portuguese verbs participate or have participated in the double past participle phenomenon. For this reason, although we can say that verbs in **C2** and **C3** are more likely than verbs in **C1** to have inherited an irregular past participle and to innovate a regular past participle, it would be wrong to claim or imply that all verbs in **C2** and **C3** that have inherited an irregular past participle will innovate a regular past participle. For example, the **C2** verb *escrever* 'write' inherited from Latin an irregular past participle *escrito* 'written', but there is no innovated regular past participle (**escrevido*).[91]

Along the same lines, our model should not imply that all verbs in **C1** which inherited a regular past participle are likely to innovate an irregular one, even though in Brazilian Portuguese we do have evidence for the extension of irregular (athematic) past participles to verbs which used not to have forms of this type (section 5.1.5).

In fact, the innovation of new forms in only a small number of Portuguese verbs is evidence that lexical specification plays a major role in the development of double past participle forms. As we saw in Chapter 4, verb-to-verb variation is one of the most important aspects of the double past participle phenomenon.

This implies that our PFM rules should not focus on changes at the level of major inflection classes **C1**, **C2** and **C3**, but instead on alterations at the level of the lexemic entry and other parts of the model where lexical data is captured. This is considered in more depth in section 5.2.2.

As further evidence that the major inflection classes are not the correct level at which to account for the development of new past participle forms, it should be noted that some diachronic

[91] This form and other similar ones may occasionally be spontaneously produced as a speech error or by children acquiring the language, but are not regularly found in spoken or written Portuguese.

trajectories affect verbs across conjugations. For example, the development of innovative irregular forms in Brazilian Portuguese, apparently by analogy with the 1SG.PRES form, is mainly attested in verbs of the first conjugation, but there are also examples of verbs in other classes undergoing the same development. An example is the second conjugation verb *trazer* 'bring' for which we have evidence of an innovated irregular form *trago*, attested alongside the inherited regular form *trazido*.[92]

A complete model of historical inflectional morphology should be able to capture the influence of many different factors on morphological change, including inflection class, although developing such a complete model is beyond the scope of this work. I focus on the major factors that play an important role in the case studies described here, arguing that integrating PFM with the LFG architecture provides an excellent theoretical platform on which fuller analyses can be built in future. The possibilities of adapting the PFM mechanisms to account for diachrony are promising, as the theory permits detailed representations that take account of both formal and functional properties.

Non-morphological factors are crucially involved in morphological and lexical change, including 'functional' factors like frequency, as well as factors from other levels of linguistic analysis, like semantics and phonology. Some of these factors are considered further in section 5.2.2, which discusses the representation of lexical specification in an LFG-PFM model of morphological change.

[92] This example is from Chagas de Souza (2015, 45). Similar examples were discussed in section 4.3.3. The Brazilian Portuguese innovative irregular past participles are discussed further in sections 5.1.5 and 5.3.3.

5.2.2 Lexical specification in an LFG-PFM model of change

This section considers the different aspects of lexical specification in relation to diachronic change, considering the different areas in which lexical information is represented in the LFG-PFM model.

The lexemic entry, which includes lexeme-level information, is the key representation of lexical information in my model. When we look at a verb which has undergone lexical change, its lexemic entry at an earlier point in time and its lexemic entry at a later point in time will be different. However, this raises an obvious question. How does this process of change occur? Relatedly, how do we represent the state of the language in which a single verb might behave in two different ways, with variation reflecting a change in progress?

A possible solution to these questions is to state that any individual speaker might have (or use) one representation of the lexemic entry of a particular verb. This coexists with another individual speaker having different lexical information in the lexemic entry of the same verb. One person's representation of a verb might differ from another person's because they have reanalysed the verb in some way during the language acquisition process.

For example, we can imagine a speaker who has not been exposed to the regular past participle form *ganhado* of the verb *ganhar* 'earn'. The speaker might innovate an irregular past participle form for this verb, based on analogy with the irregular past participle form *gasto* of the semantically and phonologically similar verb *gastar* 'spend'. When called upon to produce the past participle form of *ganhar*, our imaginary speaker would produce a new irregular form *ganho*, which has not previously been part of the language. This form might later be acquired by language learners, coming to spread through the speech community, and eventually replacing the regular form *ganhado*.

This is a model of language change based on acquisition, a concept that goes back at least as far as the 19th century. Sweet (1888) described the 'defective' acquisition of sounds by children, and

Müller ([1862] 2010, 57) described the regularisation of inflectional morphology by children, as summarised by Diessel (2012). This model was later developed in the Principles and Parameters framework (e.g. Lightfoot 1991) as well as in other theoretical approaches (e.g. Anderson 1973). Applying this model to an LFG-PFM perspective, a single speaker only has one representation of the lexemic entry for the verb, but there is variation between speakers. Thus our set of PFM rules never needs to generate alternative forms in the same context; we just need two different sets of lexemic entries (or PFM rules), one of which is used by the conservative speaker, and one by the innovative speaker.[93]

In the example just given, and similar examples of change, various different pieces of information are involved in the PFM conception of past participle formation set out in Chapter 4. The existence of an irregular stem root (stem root l) used in the formation of the irregular past participle form is one of these. This is taken to be represented in the lexemic entry for the verb (as the form of irregular stems is arbitrary, lexeme-specific information). Lexical data of this type can also be presented in an array or table of stem forms, as shown in Chapter 4, but this should be considered ancillary apparatus provided to make it easier to understand the data being presented, rather than a crucial part of the analysis. All the information contained in these tables is represented in the lexemic entry.

[93] As noted above, this is not to say that a speaker's mental grammar is fixed: it can develop throughout their lifetime. The model represented here is a simplification. A similar model can account for dialectal variation within speakers: for example, if a speaker frequently uses a standard and a non-standard variety, they could use different representations of the same lexical item depending on which variety they speak. This is not developed further here.

Another part of the PFM model that changes is the definition of the inflection classes that are specifically designed to capture the behaviour of past participle forms, such as **IrregPTCP** and **DoublePTCP**. The definition of these inflection classes, i.e. which verbs are their members, will change over time as verbs move in and out of the relevant inflection class, as discussed in section 5.3.2 below. The definitions of these classes are given by PFM rules, but they could also be considered part of the general concept of 'lexical information', since they determine which lexemes follow which morphological patterns.

The representation of lexical information in the LFG-PFM model is fairly complex, with different facets of lexical data represented in different places. This allows a thorough representation of the different relevant facts, but it can also appear piecemeal. For this reason, in section 5.3 I work through examples of how the LFG-PFM model represents diachronic change and variation, including various changes to lexical information, and in section 5.3.5, I present sample lexemic entries to demonstrate how morphological change results in different information in the lexicon.

Before moving on to look at these worked examples of the representation of diachronic change, I consider another vital aspect of the changes seen in the past participles, namely analogy. Analogical processes are a driving force behind many grammatical developments, and changes in the double past participle phenomenon are a clear example of the importance of analogy. The next section therefore looks more closely at the concept of analogy and its relevance to the Portuguese past participles, as well as possible directions for its treatment within LFG-PFM.

5.2.3 Analogical change in the diachrony of Portuguese past participles

Analogy is a crucial factor in language change in general, and we have seen throughout the chapter so far that it has had wide-reaching effects on the development of Portuguese past participles, particularly through the creation of new forms that have some resemblance to existing forms in other paradigms. Analogy can be defined as:

> 'Process by which a form *a* is either changed or created in such a way that the difference between it and a contrasting form *b* becomes the same as that of other pairs of forms whose meanings or categories contrast similarly.' (Matthews 2007)

Analogy can be seen as a kind of logic by which changes take place, rather than as a type of change itself. Morphological changes of many different kinds can proceed by analogy: in some instances, these changes promote regularisation, but analogy is also responsible for the extension of irregular patterns. Recent work on analogy has investigated constraints on what types of changes are possible or likely (Fertig 2013, chap. 7).

As we have seen, standard PFM is intended to describe synchronic systems. For this reason, there is no mechanism in PFM by which analogical processes can occur across paradigms as a language's morphology undergoes change. However, information about the various factors known to be relevant to analogy might be included in a full model of language change. For example, our model might identify similarities between verbs (in inflection class, phonology, semantics or other aspects) that are more likely to influence each other analogically. The LFG-PFM model already contains much information of this sort, so future development of the model could bring together this information to predict likely patterns of analogical change in diachrony. Several factors could be taken into consideration in such a model: semantics, phonology, inflection class, frequency and syncretism.

We saw in section 5.2.2 above that that the verbs *gastar* 'spend' and *ganhar* 'earn' both come to have irregular past participles, with *ganhar* apparently changing by analogy with *gastar*. The two verbs are from the same semantic field (to do with monetary transactions), and they have clear phonological similarity, consisting of two syllables, with the same onset at the first syllable. They are also members of the same major inflection class, **C1**. This example therefore includes three factors that could encourage a process of analogy between two verbs (in this case, the innovation of an irregular past participle): semantics, phonology and inflection class.

A further crucial factor in analogical change is frequency (Bybee 1985; Diessel 2007; Aaron Smith 2012), both type frequency, which is a measure of the frequency of the whole lexeme (Fertig 2013, 113) and to a lesser extent token frequency, the frequency of individual wordforms (Fertig 2013, 118). Type frequency is particularly relevant to the extension of morphological change by analogy: higher type frequency patterns are more likely to be extended. Items with low token frequency are more likely to regularise by analogical processes, but even very high-frequency items are affected by other (non-regularising) analogical changes (Fertig 2013, 132).

The relevance of frequency has already been explored at the synchronic level in LFG. When two or more forms exist and are in competition (variation), the likelihood of one or another being used depends partly on the relative frequency of the forms. This type of probability at a single point of time has been extensively studied within LFG using computational models which combine the theoretical grammar of LFG with stochastic methods (Johnson 2000; Hautli and King 2009; Forst 2011; Hautli, Çetinoğlu, and van Genabith 2010).

Future work on analogy could incorporate these insights about frequency and probability into a LFG-PFM model of change. The parallel architecture of LFG with the integration of a PFM component as discussed here would provide an ideal platform to investigate these questions, since different interacting levels of grammatical structure can be captured in this model, and functional properties such as frequency can be included and represented, following from the stochastic work cited above. For example, Johnson (2000) shows how an LFG model can be supplemented with a stochastic component which represents the likelihood of particular forms being selected, based on probability estimates reflecting frequencies from a training corpus.

Returning to the case study at hand, syncretism has also been argued to be relevant to analogical change. In the extension of irregular past participles in the Brazilian innovative forms, Chagas de Souza (2015) and Schwenter et al. (2019) argue that the existence of a syncretism in some verbs, between their irregular past participle and the first person singular present indicative, predisposes

other verbs to undergo the analogical change that creates the irregular past participle form. Schwenter et al. (2019) show that the most frequently used irregular past participles are those that coincide with the first person singular present indicative, suggesting that this syncretism reinforces the analogical process. The syncretism is represented in my PFM rules through the extension of a particular stem root (k, usually restricted to the first person singular present indicative) to the formation of the past participle. Representations of syncretism in PFM could also be included in a wider LFG-PFM model of analogical change, if it is the case that syncretic patterns are a relevant factor in analogical processes.

This section has suggested future directions for a full account of analogical processes in the LFG-PFM model used here. The parallel architecture of LFG, which includes multiple distinct grammatical modules, enables the inclusion of information about the many factors that are relevant to analogy, particularly when integrated with the powerful morphological model of PFM. These factors could include inflection class, semantics and phonology, as well as syncretism within the paradigm, and frequency.

Using the LFG-PFM model described in this thesis, we can develop rules to account for forms that have undergone analogical change, without necessarily capturing why these particular analogical changes have taken place. The next section demonstrates that the PFM rules proposed for Portuguese past participles in Chapter 4 can be adjusted to account for the diachronic divergence and regional variation seen throughout the history of Portuguese.

5.3 PFM rules to account for diachrony

As we have seen in the previous sections of this chapter, there are complex patterns of diachronic development in the double past participle phenomenon, leading to divergence across varieties of Portuguese. This section presents a PFM-based model of these changes, situated within the LFG architecture, which demonstrates how different outcomes are represented in the morphology and the lexicon. As part of this presentation, we see how the LFG-PFM model can represent patterns

of morphological variation at a single point of time between speakers and between varieties, with morphological changes represented in verbs' lexemic entries, while other aspects of syntax are unaffected by morphological change. Variation is captured here in a series of static analyses that together represent the relevant incremental differences between the grammars of different speakers' language varieties.

Lexical specification plays a significant role in the double past participle phenomenon, as demonstrated in section 5.2.2. This theme is taken up in section 5.3.1. As we saw in Chapter 4, the double past participle phenomenon is essentially a lexical phenomenon, in that there is arbitrary variation between verbs as to whether they participate in the phenomenon. This specification is represented at a lexical level. In my model, the same verb at different points in time will have differences in the PFM lexical data and LFG lexemic entry.

Section 5.3.2 considers the role of inflection classes and their interactions with stem formation rules in accounting for double past participles, showing that the existing rules can be adjusted to provide fine-grained changes in the representation of verbs' different forms. Section 5.3.3 builds on the rules proposed so far, adapting them to account for areas where Brazilian Portuguese and European Portuguese diverge, namely the existence of the Brazilian innovative past participles and the loss of the active/passive distinction in some Brazilian Portuguese varieties. I then consider the boundaries of representing change in the PFM model (section 5.3.4), looking at possible directions for representing the movement of forms across paradigms in suppletion. Finally, section 5.3.5 gives sample lexemic entries to demonstrate how they reflect changes in morphological properties.

5.3.1 Stem roots

In this section, I discuss how the diachronic development of the past participles relates to stem roots, as listed for each verb in the PFM lexical data.

As we saw in Chapter 4, in my account of modern European Portuguese data, verbs with an irregular past participle often have an irregular basic stem root, namely stem root l (abbreviated sr_l). For example, the verb *aberto* has the irregular basic stem root l 'aber'. This information is listed in the lexical data for modern Portuguese verbs, given in Appendix III and repeated here as Table 33 (p240) for convenience.[94]

Verbs with a lexically specified basic stem root l always have an irregular past participle. For verbs with no specified stem root l, the past participle (whether regular or irregular) is formed from the basic stem root g (abbreviated sr_g). Modern European Portuguese verbs with an irregular past participle that is formed from stem root g include first conjugation verbs like *gastar* 'spend', whose past participle is *gasto*, a truncated form.

Here I discuss how some of the various diachronic trajectories relate to stem root g and stem root l in my PFM analysis, looking initially at verbs in the first conjugation class (**C1**) before considering verbs in the second and third conjugation classes (**C2** and **C3**).

[94] In the table, verbs below the dotted line (*abrir* 'open' and below) are those which are not analysed by Taylor (2012).

	English translation	**sr$_g$**	**sr$_h$**	**sr$_k$**	**sr$_l$**
CANTAR	sing	kant	—	—	—
COMER	eat	kom	—	—	—
PARTIR	break	part	—	—	—
CABER	fit	kab	kob	kaib	—
CRER	believe	kr	—	krei	—
DAR	give	d	—	—	—
DIZER	say	diz	dis	dig	di
ESTAR	be	est	estiv	esteZ	—
FAZER	do	faz	fiz	fas	fei
LER	read	l	—	lei	—
MEDIR	measure	med	—	mes	—
OUVIR	hear	ov	—	os, oj	—
PEDIR	ask	ped	—	pes	—
PERDER	lose	perd	—	perk	—
PODER	be able	pod	pud	pos	—
QUERER	want	ker	kis	keir	—
RIR	laugh	r	—	ri	—
SABER	know	sab	sob	saib	—
TRAZER	bring	traz	troS	trag	—
VALER	be worth	val	—	vaL	—
ABRIR	open	abr	—	—	aber
ACEITAR	accept	aseit	—	—	—
ACENDER	light, turn on	asend	—	—	ase
COBRIR	cover	kobr	—	—	kober
ELEGER	choose	eleZ	—	—	elei
ENTREGAR	hand over	entreg	—	—	—
ESCREVER	write	eskrev	—	—	eskri
EXPRIMIR	express	esprim	—	—	espres
GANHAR	earn	gaɲ	—	—	—
GASTAR	spend	gast	—	—	—
LIMPAR	clean	limp	—	—	—
MORRER	die	morr	—	—	mor
PAGAR	pay	pag	—	—	—
PÔR	put	p	—	poɲ	pos
PRENDER	arrest	prend	—	—	pre
SUSPENDER	suspend	suspend	—	—	suspen
VER	see	v	—	veZ	vis

Table 33 Basic stem roots of modern Portuguese verbs

5.3.1.1 First conjugation verbs

First conjugation class verbs with irregular past participles verbs generally inherited a regular past participle form from Latin, as described in section 5.1.3.1 above. At some point in their history

some of these Portuguese verbs have come to have irregular past participles which are athematic or truncated in form: these are formed from basic stem root g. These verbs do not have a lexically specified stem root l. These irregular past participles take the form:

Stem root g + o

By contrast, the regular past participles of these verbs take the form:

Stem root g +a +d +o

There is only one exception to this: the verb *matar* 'kill' has the irregular past participle form *morto*, which has come from the paradigm of *morrer* 'die' as a suppletive form. This exceptional case is considered separately in section 5.3.4, where I discuss capturing diachronic change in paradigm membership (through reanalysis) in PFM.

Changes in the basic stem roots are thus not needed to account for the development and extension of irregular past participles of verbs in **C1**.

5.3.1.2 Second and third conjugation verbs

Verbs in the second and third conjugation classes may have inherited from Latin an irregular (athematic) past participle. In my analysis, throughout their history as Portuguese verbs, these verbs have had a lexically specified (unpredictable) stem root l.

The irregular past participle of these verbs in Portuguese has a complex stem (represented by **ɛ)** formed on the basis of stem root l. Examples of second and third conjugation class verbs with irregular past participles include *acender* 'light, turn on' (irregular past participle *aceso* from Latin *accesus*) as well as others detailed above (*eleger* 'choose', *exprimir* 'express', *morrer* 'die', *prender* 'arrest', *suspender* 'suspend'). Some of these verbs have only this irregular past participle, such as *escrever* 'write' (past participle *escrito*) and *abrir* 'open' (past participle *aberto*).

By contrast, some second and third conjugation class verbs that inherited irregular past participles from Latin (with complex stem ɛ based on stem root l) also developed new regular past participles. The regular past participle, with complex stem c, is formed from the regular basic stem root g, as is the case for *acender* 'light, turn on' (past participles *aceso* and *acendido*). These regular forms arose by analogy with other verbs' regular past participle forms.[95]

If both irregular and regular forms remain in use, these verbs are in the **DoublePTCP** class. Alternatively, their irregular past participle can fall out of use. By this mechanism, verbs in **C2** or **C3** which inherited an irregular past participle from Latin, with stem root l, can eventually undergo a full regularisation: first a regular past participle form is created (on the basis of stem root g), then this regular form becomes more frequent, and the irregular past participle form (with stem root l) falls out of use. This leads to stem root l being deleted from the lexical data for this verb. Previously irregular verbs thus come to follow the more widely applicable rules.

Verbs which have undergone this diachronic change were described in section 5.1.3.2. Examples given by Said Ali ([1908] 2008, 155–56) include the loss of the irregular past participle forms[96] *cinto* from *cingir* 'gird' *(cingido), nado* from *nascer* 'be born' *(nascido), resoluto* from *resolver* 'resolve' *(resolvido)*, *diviso* from *dividir* 'divide' (*diviso*) and *defeso* from *defender* 'defend' (*defendido*).[97] At one point these

[95] See section 5.2.3 for a discussion of analogy and a brief exploration of the difficulties of representing it in PFM.

[96] Regular forms, now the only surviving past participles of these verbs, are given in parentheses.

[97] As an interesting aside, Ali claims that the regular forms of these verbs became the preferred form relatively recently, while there are also examples of verbs where the regular form is preferred throughout the history of Portuguese, despite the Latin verb having only an irregular form: the Portuguese forms *movido* 'moved', *metido* (or *metudo*) 'placed', *recebido* (or *recebudo*) 'received' and

verbs would have had a specified stem root l[98] to account for their irregular forms inherited from Latin, but these stem roots have subsequently fallen out of use, with the regular past participle form (formed on the basis of stem root g) surviving.

The presence or absence of stem root l in the lexical data thus reflects the existence of an irregular past participle, which is reflected in my diachronic account.

5.3.1.3 Summary

This section has shown that the basic stem roots listed in the lexical data reflect one aspect of possible morphological change. In the process of diachronic development from Latin to modern Portuguese, new irregular basic stem roots for the past participle are not created. There are no examples, for example, of verbs in the second or third conjugation classes innovating an irregular stem root l along the lines of the stem root l that other verbs in the classes have, such as */komi/ giving the past participle **comeito*. Verbs may retain their irregular stem roots, or lose them (regularising by analogy), but do not gain new stem roots.[99]

---

vencido (or *vençudo*) 'won, conquered' were never in competition with irregular forms that would be expected from the Latin participles *motum*, *missum*, *receptum* and *victum* (Ali [1908] 2008, 156). It would be interesting to trace the details of the chronology of this regularisation, i.e. whether it occurred in Late Latin or in some unattested form of Portuguese, but this is beyond the scope of this thesis.

[98] Namely: 'na', 'resolu', 'divi', 'defe'

[99] Again, the suppletive inclusion of *morto* in the paradigm of *matar* 'kill' is an exception.

As we shall see in section 5.3.3, when new irregular past participles are formed (such as the Brazilian Portuguese innovative past participles), they are based on a stem that is already in the verb's lexical data.

Having seen in this section the relevance of stem roots to diachronic change in the double past participle phenomenon, I now move on to consider inflection classes and stem formation rules.

5.3.2 Stem formation rules and inflection classes

In my PFM analysis, participation in the double past participle phenomenon is conditioned by membership of particular inflection classes, such as the classes **DoublePTCP** and **IrregPTCP**. In this section I discuss the effects of changes in membership of these inflection classes and their interaction with the stem formation rules, in relation to the diachronic development of double past participles.

In my rules for modern European Portuguese, the regular past participle (complex stem c) is formed through the Block 0 stem formation rule 122:

122. BSR_0,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[stemg[6]], ι>

The irregular past participle (complex stem ɛ) is formed through the Block 0 stem formation rule 123, which selects stem l (as discussed above) and the null theme vowel [9]:

123. BSR_0,{ɛ},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[steml[9]], ι>

Where stem l is not specified in the lexical data, it is reindexed to stem g: this means that for verbs with no stem l specified, the more basic stem g is used (see section 3.3.1 for further details on reindexing).

The output of the Block 0 rules then feeds into the Block 1 stem formation rules. In Block 1, different rules apply according to the verb's membership of various inflection classes that account

for the formation of irregular past participles, namely **IrregPTCP** and its subclasses **TruncPTCP, T-IrregPTCP** and **S-IrregPTCP**.

The default rule, which applies to verbs which are not in the class **IrregPTCP**, provides the consonant /d/:

126. SR_1,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/d/], ι>for all i

This rule is not particularly relevant to diachronic change or synchronic variation, except that it may come to apply to more verbs as they regularise by analogy, exiting the class **IrregPTCP**.

5.3.2.1 Truncated-type past participles

One subclass of verbs in the **IrregPTCP** class has a truncated past participle. In standard Portuguese, all of these verbs are in the first conjugation class **C1**. To form the past participle, rule 129 applies, which does not insert a consonant.

129. SR_1,{ɐ},**TruncPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi], ι>for all i

The irregular past participle is therefore differentiated from a regular past participle by its null theme vowel and lack of consonant (e.g. *gasto* in place of *gastado* for *gastar* 'spend'). This is quite different from the formation of the irregular past participle of **C2** and **C3** verbs, which belong to one of the subclasses **T-IrregPTCP** and **S-IrregPTCP** (section 5.3.2.3).

The question arises of how diachronically these verbs come to have an irregular past participle form at all during the history of Portuguese. We saw above that a process of analogy with other similar forms, perhaps nominal forms (rather than past participles), has been suggested to explain the creation of the form *pago* which could then have acted as a model for the formation of other irregular past participles like *gasto* 'spend' and *ganho* 'earn'.

In my model, for this analogical extension to apply, a verb would immediately come to have the correct irregular form simply by becoming a member of the inflection class **TruncPTCP**. This is,

however, a simple restatement of the facts, rather than an explanation: the PFM model cannot explain *why* such an analogical change would take place, because analogy is unpredictable: there is no principled way of predicting when a lexeme will undergo analogical change, even though certain factors are known to be relevant (see section 5.2.3). On the other hand, formulating and comparing PFM models allows a more detailed understanding of *how* these changes proceed within or across morphological systems.

5.3.2.2 Verbs in the subclasses T-IrregPTCP and S-IrregPTCP

Verbs in the second and third conjugation class that have irregular past participles belong to either the **T-IrregPTCP** or the **S-IrregPTCP** inflection classes in the standard variety. Their irregular past participles take the form:

Stem root l +{s|t} +o

Their regular past participles take the form:

Stem root g +i +d +o

As we saw in section 5.1.2, the irregular forms in these conjugation classes inherited from Latin may be 'tautic' (with consonant /t/ before the final vowel) or 'sigmatic' (with consonant /s/ before the final vowel). This is captured in the PFM rules for Portuguese by the stem formation rules 127 and 128, which provide the appropriate consonant for words in the inflection classes **T-IrregPTCP** and **S-IrregPTCP** for the irregular complex stem **ɛ**:

127. SR_1,{**ɛ**},**T-IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/t/], ι> for all i

128. SR_1,{**ɛ**},**S-IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/s/], ι> for all i

Having seen the Latin data (section 5.1.2.2), the existence of these very specific inflection classes is now clearly justified. These lexical patterns cannot be explained by reference to any synchronic

process, phonological or otherwise. Instead, they are due to inherited morphological irregularity which can be traced directly back to athematic participle forms in Latin.

By contrast, the regular forms have the consonant /d/ which is found across the three major conjugation classes (**C1**, **C2** and **C3**) after the theme vowel, and is provided by rule 126:

126. $SR_1,\{c\},\mathbf{V} \cap \mathbf{Pt}(<\chi, \iota>) =_{def} <[\chi i/d/], \iota>$ for all i

Some verbs in the same classes (**C2** and **C3**) which inherited only an athematic form from Latin never innovated a regular form, maintaining only an irregular past participle throughout their history as Portuguese verbs. This is the case for verbs like *abrir* 'open', *escrever* 'write', *cobrir* 'cover', *pôr* 'put' and *ver* 'see'. These verbs have not been subject to the same innovation of regular past participle forms as their counterparts, showing some resistance to the process of analogy.[100]

In any case, both these verbs with only irregular past participles and the double past participle verbs in **C2** and **C3** retain their membership in the same class of either **T-IrregPTCP** or **S-IrregPTCP**. This demonstrates the diachronic resilience of these inflection classes, which can be traced to Latin.

A verb could stop being a member of one of these subclasses if its irregular past participle falls completely out of use; that is to say, if it leaves the class **IrregPTCP** it will also leave the relevant subclass, **T-IrregPTCP** or **S-IrregPTCP**. I discuss below how verbs might move out of the wider inflection class, **IrregPTCP**, as part of the next section on the role of the subclasses **IrregPTCP** and **DoublePTCP** in the conditioning of the double past participle phenomenon.

---

[100] We saw in section 5.2.3 that verbs with particular frequency patterns, for example, may be disposed towards analogical processes. Other characteristics may equally make certain items resistant to these processes.

5.3.2.3 The classes DoublePTCP and IrregPTCP in the double past participle phenomenon

The class **DoublePTCP** is highly relevant to the double past participle phenomenon. This subclass of **IrregPTCP** accounts for differences in function (that is, active and passive properties) rather than purely differences in form (verb morphology). Over a period of time, a verb's membership of this class may change, reflecting changes to the function of its irregular form.

A verb stops being in the class **DoublePTCP** (but remains in **IrregPTCP**) if its regular past participle stops being used. For the sake of an example, I use the verb *gastar* 'spend'. Modern corpus data and experimental work which I carried out in Lisbon shows that the regular past participle *gastado* is no longer used in European Portuguese, even in the active context when it might be expected (Thomas 2018).

As it has a truncated irregular past participle form, *gastar* is a member of the **TruncPTCP** subclass, and therefore of the **IrregPTCP** class:

108. { ENTREGAR, GANHAR, GASTAR, LIMPAR, MATAR, PAGAR …} = **TruncPTCP** ⊂ **V** ∩ **Pt**

111. **S-IrregPTCP** ∪ **T-IrregPTCP** ∪ **TruncPTCP** = **IrregPTCP** ⊂ **V** ∩ **Pt**

At an earlier stage of Portuguese, the verb *gastar* would have also been included in the definition of the **DoublePTCP** class:

112. {ACENDER, EXPRIMIR, PRENDER, GASTAR …} = **DoublePTCP** ⊂ **V** ∩ **Pt**

The lexemic entry of the verb also includes the information that it is in these inflection classes (see section 5.3.5).

Membership of the **IrregPTCP** and **DoublePTCP** classes would mean that the following rules applied, associating the regular stem (defined as stemc [10]) with past participles with the feature

'voice: active', and the irregular stem (defined as stemɛ [8]) as the default past participle (for past participles without the feature 'voice: active').

144. SSR_0,{vform: ppl}, (**IrregPTCP**) ∩ **Pt**(<χ, σ>) $=_{def}$ <[stemɛ [8]], σ>

145. SSR_0,{vform: ppl, voice: active}, (**DoublePTCP**) ∩ **Pt**(<χ, σ>) $=_{def}$ <[stemc [10]], σ>

In the grammar of a speaker who acquired the verb *gastar* without acquiring the regular past participle form (stemc [10]), rule 145 would not apply to the verb *gastar*, and the verb would not be listed as a member of the **DoublePTCP** class, although it would still be a member of the **IrregPTCP** class.

The rule defining the **DoublePTCP** class would therefore adapted to reflect this change:

112.* {ACENDER, EXPRIMIR, PRENDER, ~~GASTAR~~ …} = **DoublePTCP** ⊂ **V** ∩ **Pt**

In my PFM model, there is no direct connection between a particular stem form and a particular feature: the connection is made by inflection class membership, allowing clear and simple adjustments to be made at the level of single lexemes, without disrupting the whole system. This is a clear advantage in accounting for lexically specified or verb-to-verb variation.

5.3.2.4 Summary

This section has shown another aspect of the PFM model's effectiveness in accounting for diachronic change. Inflection class membership is crucial to the diachronic development of the double past participle phenomenon, as it conditions the application of the stem formation rules that determine the form of the past participle. In this way, as past participles come into use or fall out of use, verbs may enter or leave an inflection class. The PFM rules also determine the form of the past participle in relation to active/passive properties according to membership of a particular inflection class. More substantial examples of possible changes to inflection classes and stem formation rules are discussed in the next section, which presents PFM rules to account for

divergences in the double past participle phenomenon in some modern Brazilian Portuguese varieties as opposed to standard European Portuguese.

5.3.3 Case studies: PFM rules for Brazilian Portuguese varieties

This section presents case studies of PFM rules to account for varieties of Portuguese other than the idealised European Portuguese model presented in Chapter 4, particularly considering the Brazilian Portuguese innovative past participle forms (section 5.3.3.1) and the loss of the active/passive distinction (section 5.3.3.2). These analyses demonstrate the usefulness of the PFM approach in accounting for incremental changes in the inflectional morphology of the language, as it changes over time. PFM inflection class definitions and other paradigm rules can be adapted independently, with an alteration in one part of the system not necessarily affecting the rest of the system, although over time one change can lead to another. This is very much in keeping with the LFG architecture, in which each level of the grammar operates separately, although with constraints between the modules. The PFM analysis given here for the Brazilian Portuguese double past participles demonstrates that a variety of divergences can be captured through different adaptations to the system of rules.

In the model presented for European Portuguese, verbs in the **DoublePTCP** class use regular forms in active contexts, and irregular forms in passive contexts. However, this connection does not hold true across all varieties of Portuguese: even for some European Portuguese speakers, it may not be the case, with the two forms of some verbs possibly being in free variation (section 4.3.2). As seen in section 5.1.5, there is little detailed evidence about the distribution of irregular and regular forms in any particular variety of Brazilian Portuguese. For some speakers, there may be no connection between active/passive contexts and regular/irregular forms respectively (Teixeira da Silva 2008). Whether these speakers have 'unusual' ideolects or represent structured or homogenous varieties of Brazilian Portuguese is open for discussion, in the absence of reliable

fine-grained studies (see section 5.1.5). Here I propose models to account for these grammars, regardless of the degree to which they are widespread.

In this presentation, new rules are marked in blue, and edited rules are numbered with a 'b' (e.g. 108b), to set them apart from the existing rule (108). Added rules are numbered with '+' (e.g. 108+) to mark where they would be inserted, and deleted rules are marked with strikethrough. The full rules proposed are given in Appendix IV, alongside the rules for European Portuguese (for a variety with the two forms split across active/passive contexts) proposed in Chapter 4.

In section 5.3.3.1 I present rules to account for the existence in Brazilian Portuguese of innovative irregular past participle forms. Examples are *pego* from *pegar* 'take, grab', *chego* from *chegar* 'arrive', *trago* from *trazer* 'bring' and *peço* from *pedir* 'ask'. New rules are required because not all of these innovative forms can be produced by the PFM rules proposed for European Portuguese irregular forms: there is evidence that the Brazilian innovative forms are built on the basis of stem root k, the stem root normally used for the formation of the first person singular present indicative (because of speakers reinterpreting the occasional identity of that form and the past participle, and extending it by analogy).

In section 5.3.3.2 I present rules for a Portuguese variety with categorical use of regular forms for some verbs (in both active and passive contexts), and categorical use of irregular forms for other verbs (again, in both active and passive contexts). These rules are compatible with the Southern Brazilian (VARSUL) data presented by Silva (2008).

5.3.3.1 Rules for innovative past participles in Brazilian Portuguese

As shown above (section 5.3.1), there is no evidence in Portuguese for verbs creating new irregular stem roots. Any new irregular past participles are formed on the basis of a stem root that is already in the verb's lexical data. This is typically the regular stem root g, which is the default basic stem root to which other stem roots (such as stem root l, used in the formation of irregular past participles) are reindexed if not specified in the lexical entry. However, evidence from innovative

past participles in Brazilian Portuguese shows that these are formed on the basis of the stem used for the formation of the first person singular present indicative (Chagas de Souza 2007; 2011; 2015). In my PFM analysis, the stem root of this verb form is stem root k, which is lexically specified for the small number of Portuguese verbs that have an irregular first person singular present indicative.

An example is the verb *trazer* 'bring' (regular past participle *trazido*). An innovative irregular past participle form *trago* is attested (154) which is identical in form to the first person singular present indicative (155):

(154) eu tinha trago meu filhote
I have.1SG.IPFV bring.INNOVPTCP my puppy
'I had brought my puppy'
(Chagas de Souza 2011, 184)

(155) As vezes trago algumas coisas pra vender
At times bring.1SG.PRES some things to sell
aqui no Brasil.
here in.the Brazil
'Sometimes I bring a few things to sell here in Brazil'
(*Corpus do Português*, Web/Dialects)

As Chagas de Souza (2015, 45) demonstrates, this type of past participle cannot be formed by deletion of the theme vowel and /d/ element: for *trazer*, this would give the unattested form **trazo*. The same process is argued to be at work in the formation of innovative past participles like *chego* from *chegar* 'arrive', which has a regular first person singular present indicative form. In my analysis, this verb has no specified stem root k listed in the lexical data (because its first person singular present indicative form is regular), so the default basic stem root g is used, following the process of reindexing (see section 3.3.1).

Chagas de Souza argues that this represents a directional syncretism[101] from the first person singular present indicative form to the past participle. In addition, he demonstrates that this syncretism has occurred by the generalisation of a coincidental analogy which already existed for some first conjugation Portuguese verbs with irregular past participles, such as *pagar* 'pay' (irregular past participle *pago*) and *gastar* 'spend' (irregular past participle *gasto*). This pattern first extended to other first conjugation verbs, like *chegar* 'arrive', where the use of the innovative past participle *chego* is described as 'non-standard' (Chagas de Souza 2011, 180), and then extended into second and third conjugation verbs like *trazer* 'bring' (*trago*).

In my model, forming innovative past participles on the basis of stem root k is effective for all the attested innovative Brazilian Portuguese past participles collected by Chagas de Souza: these include *chego* from *chegar* 'arrive', *compro* from *comprar* 'buy', *falo* from *falar* 'speak', *fico* from *ficar* 'stay', *peço* from *pedir* 'ask', *perco* from *perder* 'lose' and *trago* from *trazer* 'bring' (Chagas de Souza 2011, 184; 2015, 52).

However, we do not want to 'overapply' this pattern. Other verbs continue to have irregular past participles formed on the basis of a stem root l inherited from Latin, e.g. *aberto* from *abrir* 'open'. Further, there are some verbs with an irregular stem root k used in the first person present indicative but not extended to the past participle, such as the verb *poder* 'be able' (regular past participle *podido*). The first person singular present indicative form is *posso*, with an irregular stem root k 'pos', but this verb does not have an irregular past participle **posso*. This means we cannot propose a general rule by which past participles in this Brazilian Portuguese variety are always formed on the basis of stem root k.

[101] As shown below, this cannot be an exceptionless syncretism.

My proposed solution is to capture the similarity between the Brazilian Portuguese innovative past participles with the existing proposed class of truncated past participles (**TruncPTCP**). This class will therefore include the verbs with an innovative past participle formed on the basis of stem root k (where specified). The class also includes the first conjugation verbs with irregular past participles which were the analogical basis of this pattern, such as *pagar* 'pay'.

The class can be defined as:

108*. {ACEITAR, CHEGAR, COMPRAR, ENTREGAR, FALAR, FICAR, GANHAR, GASTAR, LIMPAR, PAGAR, PEDIR, PEGAR, PERDER, TRAZER …} = **TruncPTCP** ⊂ **V** ∩ **Pt** *edited*

As above, the class **TruncPTCP** is defined as a subset of the class **IrregPTCP**: this also applies to the Brazilian Portuguese innovative past participles, which are irregular.

111. **S-IrregPTCP** ∪ **T-IrregPTCP** ∪ **TruncPTCP** = **IrregPTCP** ⊂ **V** ∩ **Pt**

In some varieties, verbs in the **TruncPTCP** class may also be members of the **DoublePTCP** class defined in rule 112, which would allow the formation and definition of a regular past participle (complex stem c) with the feature 'voice: active'.

112. {ACENDER, PRENDER, EXPRIMIR …} = **DoublePTCP** ⊂ **V** ∩ **Pt**

However, as seen in section 5.1.5 above, some Brazilian Portuguese varieties may not retain an active/passive distinction for past participles. The voice properties of past participles in Brazilian Portuguese are discussed in the next section (5.3.3.2).

The stems listed in the lexical data for the verbs in **TruncPTCP** remain unchanged. This fits with the generalisation made above in section 5.3.1 that Portuguese verbs do not innovate new irregular stem roots. However, their new inflection class membership will be listed in the LFG lexemic

entries for these verbs, reflecting the fact that the verbs have different morphological patterns in these varieties from in European Portuguese. Example lexemic entries are given in section 5.3.5.

One important point in the development of the Brazilian Portuguese innovative past participles is the reanalysis of the stem root used in their formation as being stem root k (normally used for the first person singular present indicative forms) rather than stem g (the regular basic stem root). This is eventually visible in forms like *trago* 'bring.PTCP'. It is reasonable to suggest that this reanalysis has applied to all verbs in the **TruncPTCP** class in Brazilian Portuguese. This is achieved in the Block 0 stem formation rules, shown below.

Recall that complex stem c is the regular past participle form, and complex stem ɕ is the irregular past participle form. For the irregular past participle of verbs in the **TruncPTCP** class, we adapt the Block 0 stem formation rules accordingly:

122. BSR_0,{c},**V** ∩ **Pt**(<χ, ι>) $=_{def}$ <[stemg[6]], ι>

123. BSR_0,{ɕ},**V** ∩ **Pt**(<χ, ι>) $=_{def}$ <[steml[9]], ι>

123+. BSR_0,{ɕ}, (**TruncPTCP**) ∩ Pt(<χ, ι>) $=_{def}$ <[stemk[9]], ι> *inserted*

No adaptation is needed to the Block 1 stem formation rules. In the rules for European Portuguese, according to rule 129, no consonant is added in the formation of complex stem ɕ for verbs in the **TruncPTCP** class. This rule applies equally to the Brazilian Portuguese innovative forms.

129. SR_1,{ɕ},**TruncPTCP** ∩ **Pt**(<χ, ι>) $=_{def}$ <[χi], ι>for all i

In the European Portuguese rules, verbs in **C2** or **C3** and in **IrregPTCP** are all members of either **S-IrregPTCP** or **T-IrregPTCP** meaning that in this rule block they either receive a /s/ or a /t/, but this is not the case for Brazilian Portuguese verbs like *trazer* 'bring' (in **C2**) or *pedir* 'ask' (in **C3**). All **TruncPTCP** verbs, regardless of whether they are in **C1, C2** or **C3**, receive no consonant in

this rule block, which is captured by rule 129 above. The extension of application of this rule beyond **C1** verbs (as in European Portuguese) to include a wider class of verbs in Brazilian Portuguese is captured by the rule describing the class membership of **TruncPTCP**.

This section has demonstrated how the forms of the innovative Brazilian Portuguese past participles can be derived by adjusting the rules proposed for European Portuguese. An obvious question is how to account for changes in their function. The rules proposed in this section assume that the distribution of active/passive properties is the same for these Brazilian Portuguese verbs as for similar verbs in European Portuguese, whether or not they are in the **DoublePTCP** class. This is not appropriate for all Brazilian Portuguese varieties (if any), so the next section identifies solutions.

5.3.3.2 Rules for loss of the active/passive distinction

As discussed above (section 5.1.5), there is evidence that at least some Brazilian Portuguese speakers have no active/passive distinction in the use of regular or irregular past participle forms. This was visible in Teixeira da Silva's (2008) analysis from the VARSUL corpus, where verbs tended to have only one form attested, either regular or irregular, whether in the active or passive context.[102]

[102] Teixeira da Silva's corpus data suggest one exception to this, the verb *morrer* 'die'. The examples given include *tinha morrido* 'he/she had died' and *está morto* 'he/she is dead'. The form *morto* in this context is likely not a part of the verbal paradigm of *morrer* 'die', instead being a simple adjective. The form *morto* is also the irregular past participle of *matar* 'kill', so *foi morto* would mean 'he/she was killed'. See section 5.3.4 for further discussion of the form *morto* in relation to *morrer* 'die' and *matar* 'kill'.

For this type of variety, where verbs either have an irregular or regular past participle with no active/passive distinction, it is possible to make dramatic simplifications to the PFM rules. This is to be expected, given the general restructuring of Portuguese morphology in Brazil due to intensive language contact, sometimes characterised as 'simplification' (Lucchesi 2001; 2017, 354–55).[103] In particular, the complex stem selection rules (143-145, repeated here for convenience) can be simplified.

143. SSR_0,{vform: ppl},**V** ∩ **Pt**(<χ, σ>) $=_{def}$ <[stemc [10]], σ>

144. SSR_0,{vform: ppl}, (**IrregPTCP**) ∩ **Pt**(<χ, σ>) $=_{def}$ <[stemc [8]], σ>

145. SSR_0,{vform: ppl, voice: active}, (**DoublePTCP**) ∩ **Pt**(<χ, σ>) $=_{def}$ <[stemc [10]], σ>

These rules (143-145) form the past participle by selecting the complex stem and the final theme vowel. The rules are adjusted as follows:

143b. SSR_0,{vform: ppl},**V** ∩ **Pt**(<χ, σ>) $=_{def}$ <[stemc [8]], σ> *(edited)*

~~144b. SSR_0,{vform: ppl}, (**IrregPTCP**) ∩ **Pt**(<χ, σ>) $=_{def}$ <[stemc [8]], σ>~~

~~145b. SSR_0,{vform: ppl, voice: active}, (**DoublePTCP**) ∩ **Pt**(<χ, σ>) $=_{def}$ <[stemc [10]], σ>~~

In varieties with two past participle forms, there can be a difference between the final theme vowel in the regular compared with the irregular past participle: for example, the verb *entregar* 'hand over' has an irregular past participle *entregu<u>e</u>* and a regular past participle *entregad<u>o</u>*. In the PFM rules for

[103] See further discussion in Chapter 1. Processes of morphological simplification are of course not limited to intensive language contact. For example, similar morphological changes have occurred in the history of (European) Portuguese in the loss of the active/passive distinction for the past participles of particular verbs, as shown in section 5.1.

European Portuguese, rule 144 selects the final theme vowel [8] (which is specified as either /e/ or /o/ for each verb) for the formation of the irregular past participle, and rules 143 and 145 select the final theme vowel [10] (which is always /o/) for the regular past participle. This difference is not required in a system with only one past participle form. The final theme vowel selected must be [8] (which varies across verbs) because the form *entregue* is retained for *entregar*, as attested in the VARSUL data (Teixeira da Silva 2008, 88).[104]

Since verbs in the variety described here have one past participle form, we do not need to specify both an irregular and a regular past participle form for any verbs, and rules 145 and 146 can be deleted.

The complex stem **ɛ** is also no longer needed. Its purpose was to exist alongside complex stem **c** so that verbs could have both a regular and an irregular past participle of different forms. In the new rules, the complex stem **c** is formed differently in the stem formation rules (122-123 and 126-129) depending on the inflection class membership of the verb, as shown below. Again, the old rules are given first in black for comparison with the new rules in blue.

[104] This is the only verb attested in the VARSUL data with a final theme vowel /e/ in this position. In European Portuguese, the irregular past participle of *aceitar* 'accept' is *aceite*, but this form is standardly *aceito* in Brazilian Portuguese and attested as such in the VARSUL corpus. Some European Portuguese speakers may use innovated irregular past participles such as *empregue* from *empregar* 'use' and *encarregue* from *encarregar* 'entrust', by analogy with *entregue* and *aceite* (Ali [1908] 2008, 160; Miara 2013, 90–91; Villalva and Almeida 2004, 5). This can easily be captured in my analysis by adjusting theme vowel [8] for these verbs for the relevant speakers.

122. BSR_0,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[stemg[6]], ι>

123. BSR_0,{ɛ},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[steml[9]], ι>

126. SR_1,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/d/], ι>for all i

127. SR_1,{ɛ},**T-IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/t/], ι> for all i

128. SR_1,{ɛ},**S-IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/s/], ι> for all i

129. SR_1,{ɛ},(**TruncPTCP**) ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi], ι>for all i

122. BSR_0,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[stemg[6]], ι>

123b. BSR_0,{c},**IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[steml[9]], ι> *edited*

123+. BSR_0,{c},**TruncPTCP** ∩ Pt(<χ, ι>) =$_{def}$ <[stemk[9]], ι> *inserted*

126. SR_1,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/d/], ι>for all i

127b. SR_1,{c},**T-IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/t/], ι> for all i *edited*

128b. SR_1,{c},**S-IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/s/], ι> for all i *edited*

129b. SR_1,{c},**TruncPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi], ι>for all i *edited*

Adjustments to the rules that define inflection classes are also needed, since the class **DoublePTCP** no longer exists. First, the European Portuguese inflection class rules are repeated:

108. {ENTREGAR, GANHAR, GASTAR, LIMPAR, MATAR, PAGAR …} = **TruncPTCP** ⊂ **V** ∩ **Pt**

109. {ACENDER, EXPRIMIR, PRENDER, SUSPENDER …} = **S-IrregPTCP ⊂ V ∩ Pt**

110. {ABRIR, COBRIR, DIZER, ELEGER, ESCREVER, FAZER, MORRER, PÔR, VER …} = **T-IrregPTCP ⊂ V ∩ Pt**

111. **S-IrregPTCP ∪ T-IrregPTCP ∪ TruncPTCP = IrregPTCP ⊂ V ∩ Pt**

112. {ACENDER, PRENDER, EXPRIMIR …} = **DoublePTCP ⊂ V ∩ Pt**

The rules suggested for this variety of Brazilian Portuguese are as follows:

108b. {ACEITAR, CHEGAR, COMPRAR, ENTREGAR, FALAR, FICAR, GANHAR, GASTAR, LIMPAR, PAGAR, PEDIR, PEGAR, PERDER, TRAZER …} = **TruncPTCP ⊂ V ∩ Pt** *edited*

109. {ACENDER, EXPRIMIR, PRENDER, SUSPENDER …} = **S-IrregPTCP ⊂ V ∩ Pt**

110. {ABRIR, COBRIR, DIZER, ELEGER, ESCREVER, FAZER, MORRER, PÔR, VER …} = **T-IrregPTCP ⊂ V ∩ Pt**

111. **S-IrregPTCP ∪ T-IrregPTCP ∪ TruncPTCP = IrregPTCP ⊂ V ∩ Pt**

112. ~~{ACENDER, PRENDER, EXPRIMIR …} = DoublePTCP ⊂ V ∩ Pt~~

To sum up, these rules remove several areas of complexity from the morphological system, reflecting the loss of the double past participle phenomenon in a Brazilian Portuguese variety where each verb has only one past participle form, used across both active and passive contexts. The class **DoublePTCP** no longer exists; past participles are not specified for active or passive properties; only one past participle form is derived for each verb; the complex stem ɛ is no longer needed and is removed from the verb system. This loss of morphological complexity reflects regularisation in the verb inflection, with the PFM rules capturing this important morphological difference between European and Brazilian Portuguese varieties. A similar account might be

available for modelling wider processes of morphological restructuring found in pidgins and creoles: this is discussed briefly in the section on directions for future research (section 7.2).

5.3.4 Cross-paradigm diachronic processes: reanalysis of paradigm membership

Having seen in depth how the LFG-PFM model permits a clear representation of the morphological divergence between Brazilian Portuguese and European Portuguese, I now take a step back to consider an aspect of morphosyntactic change that is more difficult to represent in this model, namely the movement of forms from one paradigm to another. Similar to the wider question of how diachronic developments can be represented in PFM, the question of cross-paradigm change is not addressed in other developments of the PFM framework, which are concerned with synchronic matters: this holds true for PFM2 (Stump 2016), GPFM (Spencer 2013) and Taylor's (2012) extension of PFM.

A full resolution of the question of how these types of morphological change can be analysed in PFM is beyond the scope of this thesis, but I present some possible steps towards including this type of information and analysing this type of change from a PFM perspective.

I consider two brief case studies of the reanalysis of forms from one paradigm as belonging to another in this section. First, I describe the integration of the past participle forms of two different Latin verbs *accipio* and *accepto* into the paradigm of a single Portuguese verb *aceitar* 'accept'. Second, I turn to the past participle *morto* which appears in the paradigms of both *morrer* 'die' and *matar* 'kill', having entered the paradigm of *matar* as a suppletive form (Maiden 2013, 506).

The incorporation of both *aceito* and *aceitado* into the paradigm of *aceitar* can be explained relatively simply, as it may not truly involve cross-paradigm phenomena within Portuguese. As we saw above (section 5.1.3.1) the verb *aceitar* is straightforwardly derived from the Latin verb *accepto* (infinitive *acceptare*), a frequentative verb ('keep accepting') built on the supine *acceptum* of the verb *accipio* 'accept'. Without diverging into an investigation of whether Latin frequentatives should be

considered a part of the extended paradigm of the base verb, it is clear that Portuguese has no productive system for the formation of frequentative verbs. There has been no distinction between 'accept' and frequentative 'keep accepting' in the history of Portuguese. Thus the Latin histories of the two past participles possibly belonging to different verbal paradigms does not affect their status within Portuguese.

The forms *aceito* and *aceitado*, respectively derived from the past participle of *accipio* (*acceptus*) and the past participle of *accepto* (*acceptatus*), are attested in 15th century Portuguese, with *aceito* predominantly in contexts with *ser* 'be' and *aceitado* in contexts with *ter* 'have'. The form *aceitado* was thus already associated with active contexts at this time, and *aceito* with passive contexts, although with some cross-over between the two, indicating the scope for variation associated with these forms and contexts. Since both forms include the regular stem /ase/ of the Portuguese verb *aceitar* 'accept', and *aceitado* is clearly a regular past participle form, there is no need for a complicated explanation of how the short form and the long form came to co-exist: we simply need to note that throughout the attested history of this Portuguese verb, it has had an irregular past participle form (with no theme vowel /a/ or consonant /d/), as formed in the PFM stem formation rules for irregular past participles of this type. The ultimate derivation of the two past participle forms from different verbs in Latin is not relevant to their existence in Portuguese: as far as we can tell from the written record, both forms have been treated as part of the paradigm of *aceitar*, suggesting an earlier date at which the distinction between the base lexeme and the frequentative form stopped being made.

The case of *morto* is more interesting. As seen above, this past participle is derived from the Latin *mortuus*, the perfect active participle of the deponent verb *morior* 'die' (passive in form, active in meaning). In Portuguese, it is found as the irregular past participle of *matar* 'kill' as well as being frequently listed as the irregular past participle of *morrer* 'die'. Maiden (2013, 507) gives the following examples with *matar*:

(156) O cão tem matado muitos gatos.
the dog has killed many cats
'The dog has been killing many cats.'

(157) Muitos gatos foram mortos pelo cão.
many cats were killed by.the dog
'Many cats were killed by the dog.'

Whether *morto* really exists as a past participle of *morrer* 'die' is somewhat complex to illustrate, since *morrer* is an intransitive verb. The irregular past participle occurs with the auxiliary 'be',[105] which is associated with passive contexts, but intransitive verbs cannot be passivised. Where *morto* does occur with 'be' it is frequently as the past participle of 'killed', as in example (157) above (where there is an agent by-phrase). Alternatively the form can have a reading as a simple adjective 'dead':

[105] To investigate whether *morto* ever occurs in a variety of Portuguese as the irregular past participle of *morrer* with auxiliary *ter*, which might be expected in some varieties where the irregular past participle is used throughout, I inspected all 63 instances of *tinha morto* 'he/she had killed/died' in the *Corpus do Português* (Web/Dialects) (Davies 2016). All of these unambiguously had the reading 'he/she had killed', with all but one having a direct object, and the one exception clearly meaning 'he had already killed [a man]'.

(158)	Quando	os	jornalistas	chegaram	ao	templo
	when	the	journalists	arrived	to.the	temple
	o	animal	já	estava	morto	
	the	animal	already	was	dead	

'When the journalists arrived at the temple, the animal was already dead.'

Since *morrer* is intransitive, the 'be dead' interpretation will always be available in instances of the copula 'be' + *morto*. It is therefore debatable whether *morto* is truly an irregular past participle of *morrer*. Even accepting this, however, we still need an account of how *morto* originally came to be in the paradigm of *matar* 'kill', since its etymology clearly points to the Latin *morior* 'die' rather than *mactare* 'kill'.

To explain this change, it appears that a semantic reanalysis has taken place, due to the ambiguity of a context like (159):

(159)	O	rato	foi	morto
	the	mouse	was	dead

'The mouse was dead'

This type of sentence can easily be reinterpreted as meaning 'The mouse was killed', as the subject of the passive sentence is in a similar syntactic relation to the verb as the subject of the copular construction with adjective 'dead'.[106] Once it has this type of interpretation, it can be used in contexts that are more clearly passive:

[106] See section 5.1.3 for a brief discussion of the LFG PREDLINK analysis of the copular context.

(160) O rato foi morto pelo gato

the mouse was killed by.the cat

'The mouse was killed (PASS) by the cat'

At this stage, the form is clearly part of the paradigm of *matar* 'kill'. From a PFM perspective, this means that the normal way that *morto* is formed in the paradigm of *morrer* 'die' must be carried over to *matar*. This involves the addition of a suppletive (lexically specified) stem root l to *matar*:

(161) Original stem roots of *matar*:

	$\mathbf{sr_g}$	$\mathbf{sr_h}$	$\mathbf{sr_k}$	$\mathbf{sr_l}$
MATAR	mat	—	—	—

With this analysis, the verb *matar* can only have a regular past participle formed on the basis of stem root g (*matado*).

(162) New stem roots of *matar*:

	$\mathbf{sr_g}$	$\mathbf{sr_h}$	$\mathbf{sr_k}$	$\mathbf{sr_l}$
MATAR	mat	—	—	mor

The addition of the lexically specified stem root l /mor/ means an irregular past participle can now be formed for this verb. However, this is not the full story: *matar* must also now be a member of the inflection class **T-IrregPTCP** because it has a past participle with the consonant /t/. This can be achieved by changing the definition of the inflection class:

110. {ABRIR, COBRIR, DIZER, ELEGER, ESCREVER, FAZER, MATAR, MORRER, PÔR, VER …} = **T-IrregPTCP** ⊂ **V** ∩ **Pt**

Membership of the subclass **T-IrregPTCP** also leads to membership of the inflection class **IrregPTCP**, through rule 111:

111. **S-IrregPTCP** ∪ **T-IrregPTCP** ∪ **TruncPTCP** = **IrregPTCP** ⊂ **V** ∩ **Pt**

By applying these small alterations to the system, the correct suppletive past participle can be added to the paradigm of *matar* and the consequences of this addition in the inflectional system can be represented. On the other hand, this account does not capture the fact that the form *morto* is taken as a whole wordform from the paradigm of *morrer*. A fuller account of suppletion and how it arises in PFM would certainly be a worthwhile area for future consideration, but is beyond the scope of this thesis. Applying PFM to morphology in diachrony represents an exciting area for further development, and the workings of cross-paradigm phenomena such as suppletion would be a ripe area for consideration.

5.3.5 Integration with LFG: Lexemic entries

This section demonstrates how information about morphological changes, as described in earlier parts of section 5.3, are captured in the LFG side of the LFG-PFM formal analysis. As we have seen, the changes involved in the double past participle phenomenon are generally captured at the level of the lexicon, with verbs moving from one inflection class to another, and with associated changes in the array of stem roots associated with each verb.

In LFG, information about lexeme-general morphological irregularities is captured in the lexemic entry. To repeat information given in section 2.3 above, the lexemic entry takes the form of a three-place relation, as described by Dalrymple, Lowe and Mycock (p440):

(163) LE < root/idiosyncratic stem forms/features, f-description, LI >

LI is the lexemic index which acts as a unique identifier for the lexeme.

The *lexemic entry* is not to be confused with the *lexical entry*, which is specific to an inflected wordform, and is a four-place relation:

(164) $\mathcal{L}$<s-form, p-form, category, f-description>

For our purposes, the most relevant part of the lexemic entry is that characterised as 'root, idiosyncratic stem forms/features'. This is where we capture the morphological attributes associated with each verb, by listing inflection class membership and lexeme-specific properties.

Example lexemic entries for the English verbs 'print' and 'drink' take the following form (Dalrymple, Lowe, and Mycock 2019, 441):

(165) LE < {M-ROOT: print, M-ICLASS:2, M-CAT:VERB},
{(↑PRED FN) = PRINT},
PRINT1>

(166) LE < {M-ROOT: drink, M-STEM1:drank, M-STEM2:drunk, M-ICLASS:3,
M-CAT:VERB},
{(↑PRED FN) = DRINK},
DRINK1>

The relative simplicity of English verbal morphology allows the relevant information to be included with just a small number of attributes, even for a verb like 'drink' which is relatively irregular by English standards. In Portuguese, by contrast, verbal morphology is much more complex, and we can therefore expect to see a considerably longer list of attributes in the lexemic entry for irregular verbs.

Applying the model of stem roots and theme vowels suggested by Taylor and adapted as described throughout Chapter 4 and Chapter 5, we must include information about stem roots, theme vowels and inflection class membership in the lexemic entries for Portuguese verbs.

A specification of inflection class membership already provides information about the morphological properties of any verb. Many properties, such as the distribution of stem roots and the identify and distribution of theme vowels, are entirely predictable from inflection class membership, depending on the degree of irregularity demonstrated by the verb. This reduces the amount of information that must be stated in the lexemic entry, as the information is predictable from the defaults for a particular inflection class.

On the other hand, elements that must be specified in the lexemic entry include any irregular stem roots, and any unpredictable theme vowels. The lexemic entry must also include the basic stem root, on the basis of which other stem roots are generally formed. The lexemic entries of a sample of Portuguese verbs, including some with irregular past participles, will demonstrate the lexemic entry's contribution.

First, I present an example lexemic entry for the regular first conjugation Portuguese verb *cantar* 'sing':

(167) LE < {M-STEMROOTG: kant, M-ICLASS:C1, M-CAT:VERB},

{(↑PRED FN) = SING},

CANTAR1>

The only stem root required to be specified for this verb is the stem root g, which is the basic stem root. This reflects the information included in the PFM lexical data for the verb, which also only lists this one stem root for the verb. The verb also belongs to the first conjugation, a major inflection class abbreviated as C1. This is the class of verbs in *-ar*. Knowing that a verb belongs to

this major inflection class allows us to predict its theme vowels, so they do not need to be specified in the lexemic entry.

Next, I consider a verb with an irregular past participle form. The second conjugation verb *abrir* 'open' has an irregular past participle form *aberto* which uses the irregular stem root l. This verb must therefore have both the basic stem root g and an irregular stem root l specified in its lexemic entry.

Inflection class membership must also be specified for this verb. As we saw in our earlier analysis, this verb's past participle includes the addition of the consonant -t-, so it belongs to the **T-IrregPTCP** inflection class (which is a subclass of the **IrregPTCP** class), as well as the major inflection class **C2**.

(168) LE < {M-STEMROOTG: abr, M-STEMROOTL: aber, M-ICLASS:C3,
M-ICLASS:T-IRREGPTCP, M-CAT:VERB},
{(↑PRED FN) = OPEN},
ABRIR1>

The theme vowels for the verb *abrir* do not need to be specified as they are inherited fully from the conjugation class **C2**.

In some instances, a verb's theme vowels are not predictable from its inflection class membership. For example, the irregular past participle *entregue* of the verb *entregar* 'hand over' has a final theme vowel -e rather than -o as we would expect. This is represented through the attribute M-THEMEVOWEL[8]:e, as below:

(169) LE < {M-STEMROOTG: entreg, M-ICLASS:C1, M-ICLASS:TRUNCPTCP,
M-THEMEVOWEL[8]:e, M-CAT:VERB},
{(↑PRED FN) = HAND-OVER},
ENTREGAR1>

Otherwise, this verb belongs to inflection classes **C1** and **TruncPTCP** (a subclass of **IrregPTCP**), and it has no irregular stems. This captures the fact that, aside from its irregular past participle (which has an unpredictable final theme vowel and causes the verb to belong to the **TruncPTCP** and **IrregPTCP** classes), the verb *entregar* is fully regular.

As we have seen above, inflection class membership and the specification of stem roots can change over time as a verb's morphological behaviour undergoes change through reanalysis or analogy. Such changes are simply reflected in the lexemic entries by the removal or addition of the relevant information.

To give a brief example, we saw above (section 5.3.2.3) that the verb *gastar* 'spend' leaves the **DoublePTCP** inflection class. At the stage when it has two past participle forms, the verb's lexemic entry has this form:

(170) LE < {M-STEMROOTG: gast, M-ICLASS:C1, M-ICLASS:TRUNCPTCP,
M-ICLASS:DOUBLEPTCP, M-CAT:VERB},
{(↑PRED FN) = SPEND},
GASTAR1>

Like *entregar* above, the verb *gastar* is in the first conjugation class and its irregular past participle is of the truncated type, so it belongs to the class **TruncPTCP**. At this stage, it also belongs to the inflection class **DoublePTCP**.

Over time, as the regular past participle is lost, the verb is reanalysed as having only an irregular past participle. It therefore leaves the class **DoublePTCP**:

(171) LE < {M-STEMROOTG: gast, M-ICLASS:C1, M-ICLASS:TRUNCPTCP, M-CAT:VERB},
{(↑PRED FN) = SPEND},
GASTAR1>

A similar approach can be taken to changes in specifications of stem roots or theme vowels, as required.

Within the Portuguese data studied here, verbs are more likely to lose irregular stem roots and less likely to gain them, while inflection class membership seems to be relatively amenable to change in both directions, as verbs can leave or join particular inflection classes. A larger-scale study could shed further light on the likely trajectories of change, revealing tendencies for particular aspects of lexemic information to change in certain ways.

The representation of morphological information in the lexemic entry allows the necessary details about inflection class membership, stem roots and theme vowels to be available to the morphological component, in this case PFM. Such morphological information is irrelevant to the syntactic components of the LFG-PFM model, since it only affects the morphological form required to realise particular feature combinations. However, in an integrated approach to morphology, syntax and their interface, it is crucial for such information to be included at some level within the grammar, to be able to account for the morphological realisations of different lexical items. The LFG architectural approach to morphology and syntax provides a clear distinction between morphological and syntactic information, allowing all the important details to be listed for lexical items as required. As we have seen, this provides a powerful model for accounting for morphological change, appropriately integrated with the wider syntactic analysis.

5.4 Conclusions

This chapter has demonstrated that the PFM framework can be used effectively to provide analyses of diachronic morphological change, giving incrementally distinctive static representations of the morphological system at different points in time. The LFG-PFM system is well-suited to capturing complex morphosyntactic phenomena, since fine-grained adjustments can be made to different parts of the account.

As we have seen in this case study of the double past participle phenomenon in Portuguese, lexical specification can play an important role in the establishment and diachronic development of morphological patterns: lexemes may change at different rates, and we thus need to be able to make changes at the level of the individual lexeme, rather than necessarily generalising changes to all verbs. The importance of inflection classes for the application of particular stem formation rules in the PFM model also allows us to account for verbs' divergent developments when they enter or leave particular inflection classes.

Considering irregular stem roots in the PFM model has given insights into the types of changes that are attested or not attested within the history of Portuguese. For example, new irregular stem roots are not added to a verb's lexical data, except in the case of the verb *matar* which gains a suppletive stem root from *morrer*. Generally, new irregular past participles are built on stem roots that already exist in the verb's paradigm, as we saw for the Brazilian innovative forms.

To sum up, the LFG-PFM model presented here provides a principled way to account for the full range of developments seen in the history of the double past participle phenomenon in Portuguese, from the inheriting of irregular athematic forms from Latin to the creation and extension of both regular and irregular forms within Portuguese, including the innovative past participles in Brazilian Portuguese. More widely, we see the advantages of integrating PFM within the LFG architecture to provide a more complete representation of the grammar, which is crucial for providing full analyses of complex phenomena across morphology and syntax. This model allows us to account

for different modules of the grammar in separate but interconnected representations, which may vary independently but also interact with each other. In particular, the case study of the Portuguese double past participle forms clarifies the links between inflectional morphology and the lexicon, represented in LFG through lexemic and lexical entries.

In the next chapter, I move on to consider grammaticalised pronouns in Portuguese, another case study in the usefulness of the LFG-PFM approach which involves different modules of the architecture, including functional structure and semantic structure alongside the lexicon and PFM rules.

Chapter 6 Portuguese grammaticalised pronouns in diachrony

The focus of this thesis now changes from verbal to nominal phenomena, taking the case study of grammaticalised first person plural and second person pronouns in Portuguese. These show mismatches between their semantics and their syntactic behaviour, or between different aspects of syntax.

We have seen in the last two chapters that combining the LFG architecture with the PFM approach to morphology provides a powerful mechanism for analysing the Portuguese double past participle phenomenon. The last chapter in particular demonstrated the usefulness of this approach for capturing diachronic change, by presenting incremental differences between consecutive grammars, as forms are gained or lost and verbs move between inflection classes. The analysis of the case study in this chapter, which focuses on mismatches at the level of semantics and functional structure rather than morphology, continues my demonstration that the LFG-PFM approach provides valuable insights into morphosyntax at many levels, including diachronic development.

The grammaticalised pronouns studied here have many interesting grammatical properties. I focus particularly on mismatches between their semantic value and various aspects of syntactic behaviour, across the grammatical features of person, number and gender, which we see in the realisation of these features by agreement on different targets. Making a close analysis of these pronouns in the LFG-PFM formal framework gives new insights into crucial distinctions to be made in the theory, and permits a new account of the interaction of semantics, syntax and morphology in grammaticalisation.

The structure of the chapter is as follows. I begin by providing some general background for the chapter in section 6.1, describing pronouns in Portuguese (section 6.1.1) and setting out major theoretical concepts needed to analyse the Portuguese data, namely agreement (section 6.1.2) and the feature bundles INDEX and CONCORD (section 6.1.3). After this introductory section, I present

data on the use of the grammaticalised pronouns in Portuguese (section 6.2), providing evidence for the historical trajectories of the pronouns, and the grammatical features associated with them in different contexts. I then present the key components of my formal analysis of the grammaticalised pronouns in section 6.3, considering category change, the representation of syntax-semantics mismatches, and the representation of grammatical features from an LFG and PFM perspective. Section 6.4 gives worked formal LFG representations of the pronouns across different varieties of Portuguese, along with sample lexical and lexemic entries, and PFM rules for the pronouns. In section 6.5 I discuss the implications of the formal analysis and finally, in section 6.6, I present the conclusions drawn from the chapter.

6.1 Introduction to Portuguese pronouns and agreement mismatches

6.1.1 Pronouns in Portuguese

The particular pronouns studied here are the second person singular pronoun *você* 'you.SG' and plural *vocês* 'you.PL', and the first person plural pronoun *a gente* 'we'. In some Portuguese varieties these pronouns have come to replace the pronouns with these functions inherited from Latin, *tu* 'you.SG' and *vós* 'you.PL', and *nós* 'we'. The sources of the grammaticalised pronouns are third person singular noun phrases: *você* 'you' is based on the noun phrase *vossa mercê* 'your mercy', while *a gente* 'we' is based on and retains the same form as the noun phrase *a gente* meaning 'the people'.[107,108]

---

[107] *Você* was originally an honorific or polite pronoun, but in modern Portuguese varieties this is not the case. The sociolinguistic characteristics of both *você* and *a gente* are discussed in section 6.2.

[108] *Vossa* is the form inherited from Latin to be used for a plural possessor. Here it corresponds to a single addressee because of earlier plurification as a politeness strategy. Plurification (Heine and Song 2011, 600) is the use of a plural pronoun to refer to a singular person, a strategy used since

To give context to the grammaticalised pronouns, a standard analysis of the full system of Portuguese pronouns (including subject, oblique, possessive and clitic forms) is set out in Table 34 below, adapted from Raposo et al. (2013, 1:902). The grammaticalised pronouns are given in bold.

The forms labelled as carrying out the 'subject' function can also be considered to represent nominative case, and similarly for oblique (prepositional), direct object (accusative), and indirect object (dative).

Roman times, and according to Thomé-Williams (2004, 88), up to the colonisation of Brazil (1500s to 1700s).

		Stress-bearing (tonic) forms			Clitics (atonic forms)		
Person and number		**Subject**	**Oblique**[109]	**Possessive**[110]	**Direct object**	**Indirect object**	**Reflexive or reciprocal**
1SG (inherited)		*eu*	*mim*	*meu, minha, meus, minhas*	*me*		
2SG (inherited)		*tu*	*ti*	*teu, tua, teus, tuas*	*te*		
2SG (grammaticalised)		***você***	***você, si***	***seu, sua, seus, suas***	***o*** / ***a***	***lhe***	***se***
3SG (inherited)	M	*ele*	*ele, si*	*seus, suas, seus, suas*	*o*	*lhe*	*se*
	F	*ela*	*ela, si*	*seus, suas, seus, suas*	*a*		
1PL (inherited)		*nós*	*nós*	*nosso, nossa, nossos, nossas*	*nos*		
1PL (grammaticalised)		***a gente***	***nós***	***nosso, nossa, nossos, nossas***	***nos, se***		***nos, se***
2PL (inherited)		*vós*	*vós*	*vosso, vossa, vossos, vossas*	*vos*		
2PL (grammaticalised)		***vocês***	***vocês***	***vosso etc., seu etc.***	***os*** / ***as***	***lhes***	***se***
					vos		
3PL (inherited)	M	*eles*	*eles*	*seu, sua, seus, suas*	*os*	*lhes*	*se*
	F	*elas*	*elas*	*seu, sua, seus, suas*	*as*		

Table 34 Pronoun forms in Portuguese: full paradigms

There are clear complexities in the relationship between different pronominal forms in Portuguese, particularly in the links between forms and functions, and the number of morphosyntactic distinctions made for different person and number combinations. This raises questions about the extent to which pronouns can be treated as having inflectional paradigms that can be accounted for in the PFM framework, which I consider in section 6.3.2.

The grammaticalised pronouns *você, vocês* and *a gente* are a fascinating part of the Portuguese pronominal system. Raposo's table of pronoun forms (Table 34) implies that these forms fit straightforwardly into the wider pronominal system, with well-established paradigms and clear relationships between forms including possessives and clitics. In fact, there is substantial variation across Portuguese in the possessive form and clitics used with the grammaticalised pronouns (if any), which is discussed in detail in this chapter, principally in section 6.2.

In the process of becoming pronouns, the source noun phrases have undergone all the major characteristics of grammaticalisation: extension, desemanticisation ('semantic bleaching'), decategorialisation and erosion (or 'phonetic reduction') (Heine and Song 2011, 591). This grammaticalisation process is sometimes called 'pronominalisation' (C. R. dos S. Lopes and Brocardo 2016, 606). Cross-linguistically, grammaticalised pronouns are not all from noun phrases, like the Portuguese examples considered here: they may also derive from different diachronic

[109] Table 34 generally shows standard European Portuguese usage, although represents some variation. In Brazilian Portuguese, the subject/oblique forms may be used as the direct object, instead of the direct object clitics, for the 3SG/3PL forms, for the grammaticalised pronouns *você* and *a gente*, and occasionally for the 1SG (Schwenter and Silva 2003, 102).

[110] The four forms given are for a possessum that is MSG, FSG, MPL or FPL in that order.

sources, such as pronouns with different features, intensifiers, nominal concepts, spatial deixis and plurification. I refer the reader to Heine and Song (2011, 600) for details.

As they come to have the same semantic value, or reference, as pronouns, the Portuguese forms undergo a change from having the syntactic properties of canonical noun phrases to having properties associated with pronouns. Their grammatical features (PERSON and NUMBER) may also eventually come to resemble those of canonical pronouns. However, there is considerable variation in the rate at which such changes occur. At a single point in time, a pronoun may therefore show different features on different 'targets', in Corbett's terminology (1991, 5). Corbett defines agreement targets as 'the item which shows agreement', applying this to gender agreement, but here I also apply the concept of 'target' to person and number features. In section 6.1.2 I show that mismatches in the person feature can be regarded as an instance of agreement mismatch.

The relevant targets for second person pronouns and first person pronouns in Portuguese are summarised in the following examples. These include some anaphoric targets (oblique and clitic pronouns). The relationship between the antecedent and its anaphors is often treated from the perspective of 'binding', that is, accounting for the semantic relation between the anaphor and its antecedent. In an LFG approach using Discourse Representation Theory (see section 6.4.4), the binding of antecedents and their anaphors is treated using an INDEX feature at semantic structure (Dalrymple, Lowe, and Mycock 2019, 523). This can be treated as a separate type of relationship from the agreement relationship that holds between the subject and other elements, such as adjectives. However, it is still important to investigate the relationship between the person, number and gender properties of the subject and those features as marked on anaphoric pronouns, so I include them here.

Second person:

Number and gender agreement on adjectives

(172)	(Tu)	estás	contento
	2SG.SUBJ	be.2SG	happy.MSG
	You are happy		

Person and number agreement on verbs

(173)	(Tu)	estás	contento
	2SG.SUBJ	be.2SG	happy.MSG
	You are happy		

Person and number agreement on possessives

(174)	(Tu)	estás	a	levar	as	tuas	malas
	2SG.SUBJ	be.2SG	to	carry	the.FPL	2SG.POSS.FPL	bags.FPL
	You are carrying your bags						

Person and number agreement on reflexives

(175)	(Tu)	já	te	viste	ao	espelho?
	2SG.SUBJ	already	2SG.REFL	see.PAST.2SG	at.the.MSG	mirror.MSG
	Have you seen yourself in the mirror?					

Person, number and (where relevant) gender features on clitic pronouns[111]

(176)	(Tu)	disseste	que	o	Pedro	te
	2SG.SUBJ	say.PAST.2SG	that	the.MSG	Pedro	2SG.OBJ
	viu					
	see.PAST.3SG					
	You said that Pedro saw you					

Person and number features on oblique pronouns

(177)	Isto	é	para	ti
	This	be.3SG	for	2SG.OBL
	This is for you			

First person plural:

Number and gender agreement on adjectives

(178)	(Nós)	estamos	contentos
	1PL.SUBJ	be.1PL	happy.MPL
	We are happy		

Person and number agreement on verbs

(179)	(Nós)	estamos	contentos
	1PL.SUBJ	be.1PL	happy.MPL
	We are happy		

[111] Second person object pronouns (e.g. *te*) are not marked for gender, but third person object pronouns (e.g. *o, a*), which may be used with *você*, are marked for gender.

Person and number agreement on possessives

(180)	(Nós)	estamos	a	levar	as	nossas	malas
	1PL.SUBJ	be.1PL	to	carry	the.FPL	1PL.POSS.FPL	bags.FPL
	We are carrying our bags						

Person and number agreement on reflexives

(181)	(Nós)	vimo	-nos	ao	espelho
	1PL.SUBJ	see.PAST.1PL	1PL.REFL	at.the.MSG	mirror.MSG
	We saw ourselves in the mirror				

Person, number and (where relevant) gender features on clitic pronouns[112]

(182)	(Nós)	dissemos	que	o	Pedro	nos
	1PL.SUBJ	say.PAST.1PL	that	the.MSG	Pedro	1PL.OBJ
	viu					
	see.PAST.3SG					
	We said that Pedro saw us					

[112] Given its third person origins, *a gente* can take 3SG agreement, which in the case of the direct object clitic would be marked for gender as well. In practice, *a gente* does not occur with third person singular marking on coreferent objects (although it can take third person singular verb agreement, adjectives etc.). Lopes (2003, 80–81) gives historic examples of *a gente* with a third person singular possessive or oblique, but argues that in this context *a gente* must be interpreted as a noun rather than a pronoun.

Person and number features on oblique pronouns

(183) Isto é para nós
This be.3SG for 1PL.OBL
This is for us

The marking of these features on all these agreement targets at different stages in the pronouns' development, and the existence of mismatches between them, are considered throughout the rest of this chapter. The next section provides some background on agreement mismatches.

6.1.2 Agreement: mismatches, imposters and hybrids

The Portuguese grammaticalised pronouns have interesting behaviour in terms of grammatical features, including agreement mismatches. Mismatches occur when the same agreement controller shows different agreement on different targets. These mismatches can take various forms, involving all three of the grammatical features of person, number and gender. Mismatches are just one aspect of the complex behaviour associated with these nominal features, which also include the resolution of features in coordination and similar contexts: see Dalrymple, Lowe and Mycock (2019, 69–74) for further discussion from an LFG perspective.

Agreement mismatches have been the object of considerable attention. Corbett (2006, 143–75) discusses many examples from a typologically diverse range of languages. For the purpose of comparison, I limit this section to a few examples of behaviour similar to that of the grammaticalised pronouns with which I am concerned.

Following the terminology used in Corbett (1979), I introduce here the notion of 'semantic' and 'syntactic' agreement. Semantic agreement is more closely aligned with the properties of the referent in the real world. Syntactic agreement, on the other hand, is based on formal characteristics.

A commonly cited example of a mismatch is found in English nouns like *committee* (Corbett 2006, 2-3,143,155). This noun generally takes singular verb agreement, which is 'syntactic', because it corresponds with a grammatically singular feature. In some contexts and for some speakers, though, the same noun phrase may take plural verb agreement, which is said to be 'semantic', corresponding with the fact that a committee has more than one member (Corbett 2006, 2, 24). We therefore have a potential mismatch between the semantic number of the noun and its grammatical number feature.

(184) The government has.SG made a decision.

(185) The government have.PL made a decision.

The phenomenon of 'imposters', in which third person forms refer to first person or second person referents, also has relevance to the grammaticalised pronouns discussed here.[113] A speaker can refer to himself or herself – the first person – using an apparently third person form such as 'this reviewer', or can use a third person form such as 'the Baroness' to address a second person.[114] In such cases, agreement behaviour is often complex. The forms used to refer to the first or second person typically use third person verb agreement, but other targets may show other properties, such as the reflexive pronoun in this example, which can be either the first person or the third person plural form:

[113] See section 6.3.1.1 for an analysis of the grammaticalised first person plural pronoun *a gente* which uses the term 'imposter'.

[114] Portuguese allows some nominal (third person) forms to be used as forms of direct address: see sections 6.2.1.1 and 6.3.1.2.

(186) In this reply, the present authors$_i^{PL}$ attempt to defend ourselves$_i^{1PL}$/themselves$_i^{3PL}$ against the scurrilous charges which have been made.

(Collins and Postal 2012, 17)

Gender mismatch patterns are also found cross-linguistically. For example, the Serbo-Croatian noun *deca* 'children' has feminine (and singular) agreement on modifiers such as adjectives (187), but neuter (and plural) agreement with co-referential pronouns (188).

(187) Posmatrali smo ovu dobru decu.
watched.1PL AUX this.FSG good.FSG children.ACC
We watched these good children.

(188) Ona su se lepo igrala.
They.NPL AUX.3PL REFL nicely played.NPL
They played well.

(Wechsler and Zlatić 2003, 6)

Nouns that show such mismatches in agreement behaviour are described as 'hybrids', terminology that was originally associated with gender agreement, but can also be used in relation to number or person features. The term 'hybrid' can also be applied to pronouns. In a paper on the Slavonic languages, Corbett (2009) identifies the Russian honorific pronoun *vy* 'you' as a hybrid, because it refers to a singular addressee, but takes plural verb agreement, and singular agreement on some targets such as the long form adjective. He gives the following examples:

(189) vy xot-ite ...
2PL.NOM want-2PL
'You (polite) want ...'

(190) vy segodnja očen' zadumčiv-aja.
2PL.NOM today very thoughtful-(LONG.FORM)FEM.SG.NOM
'You are very thoughtful today.' (one woman addressed politely)
(Corbett 2009, 10)

In (189), the honorific form *vy* 'you' takes a second person plural verb form, even when addressing a single individual. In (190), the same form is used with a singular long-form adjective.

The term 'hybrid' can be aptly applied to the Portuguese grammaticalised pronouns discussed here, which demonstrate similar behaviour, as we see in section 6.2.

Another important concept in the study of agreement mismatches is the agreement hierarchy. Corbett (1979; 1991) demonstrates that a cross-linguistic agreement hierarchy governs gender agreement for hybrid nouns.

(191) Agreement hierarchy:
Attributive adjective > Predicative adjective > Relative pronoun > Personal pronoun

Attributive adjectives are most likely to agree with the syntactic gender of the noun. The syntactic gender may align with that expected from the morphological form: for example, the German *Mädchen* 'girl' has neuter syntactic gender, which is common for nouns ending in *-chen*. At the other end of the scale, personal pronouns are more likely to agree with the semantic gender of the noun. This would be feminine for *Mädchen*: however, both neuter and feminine pronouns are found co-referring with *Mädchen* (192).

(192) Das Mädchen ist jung/*junge;
the.NEUT girl be.3SG young.NEUT/*F
ich kenn sie/es.
I know.1SG 3SG.NEUT/F
'The girl is young; I know her.'

(Example adapted from Wagner 2003)

There is considerable cross-linguistic variation in where on the scale a given hybrid noun will switch from syntactic to semantic gender agreement. The agreement hierarchy is also diachronically relevant (Corbett 1991, 248ff): for a hybrid noun that changes over time from syntactic gender to semantic gender, the change will take place gradually along the agreement hierarchy, with co-referring personal pronouns changing first, then the relative pronoun, and so on. Corbett describes the first stages of this change occurring in Chichewa, where the noun *kamwana* 'small child' shows either (conservative) syntactic agreement or (novel) semantic agreement in emphatic pronouns, but syntactic agreement elsewhere. Corbett (2006, 218ff) also suggests that the same hierarchy is relevant for number agreement. In addition, he proposes that the Predicate Hierarchy, originally suggested by Comrie (1975), interacts with the agreement hierarchy as below (Corbett 2006, 233):

(193) Attributive adjective → Predicate verb → Rel. pronoun → Personal pronoun
↘
Participle → Adjective → Noun

The original insight of the Predicate Hierarchy was that in agreement patterns seen with polite pronouns, 'more verb-like predicates tend to agree with features of the surface subject, while more noun-like predicates tend to agree with the underlying subject'(Comrie 1975, 406). This is clearly relevant to the agreement of the grammaticalised Portuguese pronouns, as discussed for the second person pronouns in section 6.2.1.

In general, the concepts of 'semantic' and 'syntactic' agreement clearly allow useful insights into complex agreement behaviour. However, as I discuss in the next section, further distinctions have also been made within 'syntactic' agreement, particularly in proposals for the existence of two distinct feature bundles, INDEX and CONCORD.

6.1.3 Beyond semantic and syntactic agreement: INDEX and CONCORD

One of the aims of this chapter is to illustrate how complex agreement behaviour, arising from historical change, can be represented formally in the LFG-PFM framework. In earlier formal work, including in LFG, similar mismatch phenomena have been accounted for using differently specified INDEX and CONCORD features within the f-structure (Wechsler and Zlatić 2000; 2003; Mittendorf and Sadler 2005; Hristov 2013; Haug and Nikitina 2016; Wechsler 2011; Wechsler and Hahm 2011; Dalrymple, Lowe, and Mycock 2019, 69–71). INDEX and CONCORD are separate feature bundles that are both relevant to agreement. The existence of two feature bundles was originally proposed within the framework of Head-Driven Phrase Structure Grammar (HPSG) (Pollard and Sag 1994). A general characterisation of the two types of features is that CONCORD features are purely syntactic and relate to inflection class, while INDEX features are closer to the intrinsic semantic features of the referent. In this respect, there are some similarities between the concept of INDEX and CONCORD and the less formal concepts of semantic and syntactic agreement; however, as we shall see, it is important to keep the pairs of concepts separate, because semantic agreement can be distinct from both INDEX and CONCORD.

In terms of the targets they are relevant to, INDEX features are 'referential indices', affecting anaphoric agreement and subject-verb agreement. By contrast, CONCORD features are 'grammatical head features', relevant to phenomena such as adjective-noun concord (Wechsler 2011). Grammatical features such as GENDER and NUMBER may be present in both INDEX and CONCORD but specified for different values. Different agreement targets may then refer either to the INDEX GENDER feature or the CONCORD GENDER feature, for example. A classic example is the Serbo-

Croatian 'deca-type' nouns, illustrated in (187) and (188) above, repeated here for convenience as (194) and (195):

(194) Posmatrali smo ovu dobru decu.
watched.1PL AUX this.FSG good.FSG children.ACC
We watched these good children.

(195) Ona su se lepo igrala.
they.NPL AUX.3PL REFL nicely played.NPL
They played well.

(Wechsler and Zlatić 2003, 6)

These nouns are analysed as having a mismatch between INDEX and CONCORD. This is schematised in (196), including an indication of the semantics and 'declension' (which corresponds to morphological features).

(196) DECLENSION ↔ CONCORD | INDEX ↔ SEMANTICS

The CONCORD features (feminine and singular) are marked on 'determiners, modifiers and non-finite predicative phrases', while the INDEX features (neuter and plural) are marked on coreferential pronouns (Wechsler and Zlatić 2000, 816). Both sets of properties can be represented at the level of functional structure in LFG (see section 6.4).

The distinction between INDEX, CONCORD and semantic agreement (as well as declension) allows straightforward accounts of particularly complex mismatches. A typical example is the Serbo-Croatian noun *braća* 'brothers' which show three different sets of properties on different agreement targets (semantic, INDEX-based and CONCORD-based) (Wechsler and Zlatić 2000, 816). These nouns can be analysed as having a three-way split, summarised in (197).

(197) DECLENSION ↔ CONCORD | INDEX | SEMANTICS

In practice, of course, such instances are very rare; mismatches of any kind are already relatively uncommon, and they are far more likely to involve only two sets of features rather than three.

Although both INDEX and CONCORD are syntactic, in practice they are often used to analyse data that has previously been conceived of as 'syntactic' versus 'semantic' agreement. Alsina and Arsenijević (2012a; 2012b) argue that marking a distinction between INDEX and CONCORD introduces unwarranted complexity into the theory, and that only syntactic and semantic agreement are needed to account for the data described by Wechsler and Zlatić. Wechsler and Zlatić (2012) provide further evidence from Serbo-Croatian data against Alsina and Arsenijević's claims, and Hristov (2013) makes further compelling arguments in favour of retaining both INDEX and CONCORD as syntactic agreement, while also retaining the possibility of purely semantic agreement. Hristov argues that Alsina and Arsenijević's proposed solution relies on problematic assumptions about the Serbo-Croatian data, including an unjustified assumption that there are pronouns with mismatched F.SG and N.PL features, and argues that 'complexity' in itself is a misguided criticism, due to the substantial redundancy and complexity found throughout language, and because much of this complexity can be eliminated by a system of feature inheritance (see discussion of this concept in section 3.3.1). The INDEX and CONCORD distinction has subsequently been fruitfully applied to analyse many languages beyond Serbo-Croatian, including Bulgarian, English, and Welsh (Mittendorf and Sadler 2005; Hristov 2013).[115]

In previous work, CONCORD has contained the features CASE, NUMBER and GENDER, while INDEX has contained PERSON, NUMBER and GENDER. My analyses also include the feature PERSON in CONCORD, which is an innovation that I propose in order to capture the changes in Portuguese pronouns under discussion. Since this chapter is broadly concerned with mismatches in PERSON

[115] Further references are given by Dalrymple, Lowe and Mycock (2019, 69–71).

feature behaviour that cannot easily be captured in conventional representations, it is necessary to specify different PERSON features in both INDEX and CONCORD. This is discussed at greater length in section 6.3.

Having seen some of the key concepts involved in the analyses of feature mismatches in grammaticalised pronouns, I now move on to describe the Portuguese pronouns studied here. As well as presenting information about their history and sociolinguistic properties, I give details of their grammatical features as shown by their agreement on different targets.

6.2 Use of grammaticalised pronouns in Portuguese

This section provides data about the use of the different grammaticalised pronouns under consideration in the chapter, namely the second person singular *você* and plural *vocês* and the first person plural *a gente*. Throughout the section, we see a contrast between the European Portuguese system, which tends towards greater retention of the inherited forms, and the Brazilian Portuguese system, where grammaticalised pronouns are used more frequently. Data about the use of different pronouns in the history of Portuguese are taken from existing secondary sources or from online corpora including the *Corpus do Português* (Davies 2016). As well as demonstrating the different uses of the grammaticalised pronouns between different varieties of Portuguese, I illustrate changes in the grammatical features (person and number) of these forms as they grammaticalise from noun phrases to pronouns, providing a number of examples. Evidence about the forms' grammatical features comes from agreement patterns seen on agreement targets: verbs, possessives and object pronouns. This demonstrates the influence of grammaticalised pronouns on the broader morphosyntactic system of Portuguese, motivating a detailed formal model of this type of change, and providing a framework for further exploration of this issue.

The structure of the rest of this section is as follows. I give several types of information about the pronouns, considering second person (singular and plural) pronouns in section 6.2.1 and first person plural pronouns in section 6.2.2. Within each section, I provide details of the general system

in which the grammaticalised pronouns play a role, looking at the range of possibilities: for example, alongside *você*, we must consider alternative forms of direct address. Then I discuss the historical development of the grammaticalised pronoun, up to the sociolinguistic implications of its use in the present day. Finally, I demonstrate the grammatical features (PERSON and NUMBER) associated with the different pronouns. In each case we see that the pronouns have interesting mismatches between aspects of their syntax and semantics, which necessitate a fine-grained formal approach. My LFG-PFM approach allows a full account of these mismatches, as demonstrated in sections 6.3 and 6.4.

6.2.1 Data: second person você *and* vocês

Here I describe the second person singular pronoun *você* and the second person plural form *vocês*. The two forms have different sociolinguistic meanings and different use distributions, and there is also considerable variation between European Portuguese and Brazilian Portuguese, as well as within these macro-varieties. For example, in many Brazilian Portuguese varieties, *você* has entirely replaced the inherited pronoun *tu*, whereas *tu* is retained in European Portuguese as a familiar form, with *você* occupying a more marginal position in the pronoun system. Section 6.2.1.1 provides more details on this. Section 6.2.1.2 describes the historical development and modern sociolinguistics of *você* and *vocês* compared with the inherited pronouns *tu* and *vós*, while sections 6.2.1.3 and 6.2.1.4 illustrate the grammatical features of *você* and *vocês*, providing evidence for agreement patterns on different types of targets, particularly looking at verb agreement, coreferent pronouns such as reflexives, and possessives.

6.2.1.1 Forms of address in Portuguese

The system of second person address in Portuguese is complex. There is substantial variation across varieties of Portuguese, both between European Portuguese and Brazilian Portuguese and within more restricted varieties. Possible strategies for the second person singular are given in Table 35.

Subject form	**Agreement patterns (verbs, oblique pronouns, possessives)**	**Historical origin**
Pronoun *tu*	- Second person (standard) - Third person verb forms (some Brazilian varieties)[116]	Inherited from Latin *tu* 2SG
Grammaticalised pronoun *você*	- Third person (standard) - Second person or mixed agreement (some Brazilian varieties)	Grammaticalised from nominal expression *vossa mercê* 'your mercy'
'Avoidance' strategy: Zero subject[117]	Third person	
Nominal strategy: With name: *o Filipe* 'Filipe', *a Ana* 'Ana' With kinship term: *o pai* 'the father', *a mãe* 'the mother' With other noun: *o senhor* 'sir' (lit. 'the gentleman'), *a senhora* 'madam' (lit. 'the lady') With title: *o professor* 'the teacher', *a senhora doutora* 'the doctor' With ceremonial title: *Vossa Excelência* 'Your Excellency', *Vossa Majestade* 'Your Majesty'	Third person	Definite article + noun

Table 35 Second person singular forms of address

[116] This pattern (e.g. *tu fala*, 2SG.SUBJ speak.3SG, 'you speak') appears to present an interesting mismatch between semantic and syntactic agreement. However, the use of the inherited second person pronoun with third person verb morphology is associated with widespread reductions in the verbal paradigm, including the levelling of all person and number combinations to use the form previously used for the third person singular (Lara Bermejo 2015, 83), creating widespread

For the second person plural, there is a similar range of possible address strategies.

syncretism. The use of *tu* with the 3SG verb therefore represents a different pathway in development from the grammaticalised pronouns *você(s)* and *a gente*. Since my focus here is on the grammaticalised pronouns, I do not consider this further here, but note that this process of interacting morphological and syntactic change represents an interesting direction for future research.

[117] An example of a Portuguese sentence with a zero subject as addressee is:

Quer *alguma coisa*?

want.3SG some thing

Do you want something?'

Subject form	**Agreement patterns (verbs, oblique pronouns, possessives)**	**Historical origin**
Pronoun *vós*	Second person plural	Inherited from Latin *vos* 2PL
Grammaticalised pronoun *vocês*	- Third person plural - Sometimes used with inherited second person plural possessive *vosso*	Plural form of *você,* grammaticalised from nominal expression *vossa mercê* 'your mercy'
'Avoidance' strategy: Zero subject	- Third person plural - Sometimes used with inherited second person plural possessive *vosso*	
Nominal strategy: With kinship term: e.g. *os pais* 'the parents' With other noun: *os senhores* 'sirs' (lit. 'the gentlemen'), *as senhoras* lit. 'the ladies' With title: *os professores* 'the teachers', *as senhoras doutoras* 'the doctors' With ceremonial title: *Vossas Excelências* 'Your Excellencies'	- Third person plural - Sometimes used with inherited second person plural possessive *vosso*	Definite article + noun

Table 36 Second person plural forms of address

The strategies have different sociolinguistic values and different patterns of use between different varieties of Portuguese. The following table aims to give a basic summary of these different possibilities, although the full details are orthogonal to the question of the agreement patterns used, which is the focus of this chapter. In the table, 'N' represents 'neutral', a term introduced by Cook (1997) to enable a more comprehensive description of Portuguese forms of address, which do not fit straightforwardly into the traditional T/V frame of analysis used by Brown and Gilman (1960).

	Sociolinguistic value in European Portuguese	**Sociolinguistic value in Brazilian Portuguese**	**Further observations**
Pronoun *tu*	T form (indicates solidarity, intimacy, camaraderie) (Cook 2013, 270)	Some varieties: not used at all Other varieties: T form.	
Grammaticalised pronoun *você*	Rarely used. Some 'negative overtones' now being lost, with *você* moving towards being an N form (Cook 2013, 286)	T form or N form	Some Brazilian Portuguese varieties have *você* with second person oblique pronouns/possessives, bringing it closer to a T form (Cook 2013, 280)
'Avoidance' strategy: Zero subject	N form (avoids making explicit claims about the relationship)	Not generally used, as subject pronouns are obligatory	
Nominal strategy: With name: *o Filipe* 'Filipe', *a Ana* 'Ana' With kinship term: *o pai* 'the father', *a mãe* 'the mother' With other noun: *o senhor* 'sir' (lit. 'the gentleman), *a senhora* 'madam' (lit. 'the lady') With title: *o professor* 'the teacher', *a senhora doutora* 'the doctor' With ceremonial title: *Vossa Excelência* 'Your Excellency', *Vossa Majestade* 'Your Majesty'	The different nominal strategies range in sociolinguistic value from N form (with name or kinship form) to V form (with titles) They are presented here with N forms higher and the most V forms lower	N or V form	

Table 37 Sociolinguistic implications of second person singular forms of address

Table 38 gives the equivalent information about the second person plural forms. The plural forms have been much less studied than the singular (Morgan and Schwenter 2019). In general, less

variation is seen in the possibilities for plural address, and there is not a straightforward symmetry between the distinctions made in the singular and those made in the plural. Morgan and Schwenter (2019, slide 58) suggest that this could be because 'interlocutor face threats are much riskier in the singular than in the plural, where address is diffused across interlocutors'.

Subject form	**Sociolinguistic value in European Portuguese**	**Sociolinguistic value in Brazilian Portuguese**	**Further observations**
Pronoun *vós*	T form, rarely used or 'obsolescing' (Morgan and Schwenter 2019)	Only retained in restricted circumstances e.g. liturgy	
Grammaticalised pronoun *vocês*	T form or N form	T form or N form	
'Avoidance' strategy: Zero subject	N form, with some formal (V) properties	Not generally used, as subject pronouns are obligatory	
Nominal strategy: With kinship term: 'e.g. *os pais* 'the parents' With other noun: *os senhores* 'sirs' (lit. 'the gentlemen'), *as senhoras* lit. 'the ladies' With title: *os professores* 'the teachers', *as senhoras doutoras* 'the doctors' With ceremonial title: *Vossas Excelências* 'Your Excellencies'	Those higher on the list are more typical N forms, those lower are V forms	N form or V form	This strategy is less commonly used for the plural than the singular, at least in European Portuguese (Amaral 2019, 59). The zero subject option is preferred.

Table 38 Sociolinguistic implications of second person plural forms of address

I refer the reader to Cook (2013) for a fuller explanation of the sociolinguistic meanings associated with the different uses. Cook's proposal that Portuguese varieties have a 'neutral' (or 'N') form of address alongside the 'T' and 'V' strategies allows a more fine-grained approach to the different address strategies, including the nominal forms listed in Table 35 to Table 38 sometimes called

'descriptive pronouns' (Amaral 2019). Following a proposal by Pountain (2003, 159) that the existence of nominal address forms 'mitigates against the survival of the original pronominal forms', Cook suggests that the existence of the nominal-type forms of address opens up many possibilities within the address system, which can account for 'a wide spectrum of sociolinguistic preferences' (Cook 2013, 288–89). The grammatical status of nominal forms of address is discussed in section 6.3.1.2.

6.2.1.2 Historical development and modern sociolinguistics of você *and* vocês

The origins of the form *você* can be traced back to the noun phrase *vossa mercê*. This noun phrase can be glossed as below:

(198) vossa mercê
2PL.POSS.FSG mercy.FSG
your mercy

Addressing someone by a noun phrase identifying an attribute of them (such as 'your mercy' or 'your excellency') was possible in many European courts in the Middle Ages, including in Portugal (Thomé-Williams 2004). Some of these such as *vossa eminência* 'your eminence' (used to address cardinals) persist as ceremonial titles to the present day in Portuguese (Thomé-Williams 2004, 87). Jensen (1981, 55) refers to these as 'medieval titles'. The case of *vossa mercê* 'your mercy' is somewhat unique, since it has undergone an extended process of grammaticalisation to become a widely used pronoun. According to Thomé-Williams (2004, 87), the address form *vossa mercê* came to be extended from addressing only the king to apply to nobles, then to the bourgeoisie, and then more widely still (a process of extension or 'vulgarization'), while other forms were created to take the place of *vossa mercê* in addressing the king. The strong sociolinguistic implications of *vossa mercê* were weakened in its grammaticalisation to *você*, as it become a general form of address with some residual meanings of politeness, which have now been entirely lost in many varieties of Portuguese.

The process by which a polite pronoun's range of use is extended, thus losing pragmatic value, has been called 'devaluation' in relation to Portuguese (Heine and Song 2011, 606), and 'pragmatic depreciation' in relation to Japanese (Ishiyama 2019, 38). A similar process has been described for English and Dutch, which have lost their inherited T pronoun in favour of a more extended V form, arguably partly due to a 'preference for V [which] then caused the group of V-receivers and V-domains to gradually grow'(Aalberse and Stoop 2015, 192). In Portuguese, the process of pragmatic depreciation of the address form went alongside the process of phonetic reduction, through the intermediary stages of *vossemecê*, *vosmecê*, and *voacê* or *vomcê* (Thomé-Williams 2004, 88).

In modern European Portuguese, Cook (2013, 286) mentions that *você* in the singular has some unspecified 'negative overtones', reflecting Cunha and Cintra's (1990, 294) statement that in European Portuguese the pronoun is only used for 'equal treatment or from superior to inferior (in age, in social class, in hierarchy) … [and] exceptionally – among certain high social classes – it appears used as an affectionate form of intimacy' [my translation]. This use 'from superior to inferior' can easily be seen to have negative overtones. According to Thomé-Williams (2004, 98) *você* is used in Portugal for unfamiliar addressees who are equal or inferior in social status and age. It is used 'to equalize persons, without showing intimacy'. It can be used in advertisements addressed to the general public, but is avoided in general social interactions, presumably due to the unfavourable implications about the addressee's social status. In Brazil, on the other hand, *você* has none of these unfavourable associations. It is the generally used form of address, unless a speaker wants to show particular respect, where a nominal strategy would be used, or (in some varieties) particular intimacy, where *tu* can be used. Cook (2013, 287–88) suggests that *você* can be seen as an 'uncommitted' form, similar to English *you*, particularly in Brazil.

As shown in section 6.2.1.1, the plural form *vocês* has a very extensive use in both European and Brazilian Portuguese, generally taking the place of the inherited second person plural form *vós*. This form is often stated to have fallen out of use in everyday language, only being retained in very

formal or formulaic contexts (Jensen 1981, 54; Cook 2013, 272). However, it is widely recognised to be retained to a much greater extent in some European Portuguese varieties (Lara Bermejo 2015; Bechara 2015, 173; Morgan and Schwenter 2019), particularly in the north of Portugal, with the full range of associated forms (such as possessive *vosso*) along with second person plural verb inflections.

Alongside the grammaticalised pronoun *vocês*, speakers may also use one of the nominal forms such as *os senhores* 'sirs', lit. 'the gentlemen', to address a group of speakers with the V form, although according to Amaral (2019, 59) this is less common than the use of the equivalent singular form in European Portuguese. In both Brazilian and European Portuguese, speakers are more likely to use a plural T form (*vós* or *vocês*) or an avoidance strategy to address a group, even if they would have addressed each of the speakers individually with the V form: Morgan and Schwenter (2019) describe this as 'asymmetry in second person plural reference', which they claim to be a common feature across Romance varieties.

The historical development of the form *vocês* has been relatively understudied (Lara Bermejo 2015; Lara Bermejo and Guilherme 2018), reflecting a more general lack of work on forms of address for the second person plural compared with the second person singular (Morgan and Schwenter 2019). In particular, there is little evidence for the development of the form from the plural source noun phrase *vossas mercês* 'your mercies'; it is likely that the plural grammaticalised form *vocês* developed in parallel with the singular form *você*, rather than being a later pluralisation of the singular form.

Despite the general lack of attention given to *vocês*, Lara Bermejo (2015) presents a detailed study of changes in forms of address for the second person plural in Portugal in the last three hundred years. He traces the gradual replacement of the inherited informal *vós* form by the (previously more formal) grammaticalised form *vocês*, with noun phrases like *os senhores*/*as senhoras* 'the gentlemen/the ladies' or kinship terms coming to fill the formal role. Lara Bermejo identifies several stages in

these processes, with geographic variation arising from differences in the speed of diffusion. His work focuses mainly on the sociolinguistic factors affecting the trajectories of different forms of address.

By contrast, my focus in this chapter is on the grammatical features of the singular *você* and its plural equivalent *vocês*, and the mismatches that arise in their use in different Portuguese varieties. I therefore now move on to consider in more detail the possible patterns of agreement demonstrated by these pronouns.

6.2.1.3 Grammatical features of você

We have seen above that the sociolinguistics of *você* is strikingly different between European and Brazilian Portuguese. There are also differences in its grammatical properties, between the two varieties, and within Brazil. We see several distinct possibilities in the grammatical features and agreement properties of *você*, all of which must be accounted for. In this section, and in the equivalent sections for *vocês* (section 6.2.1.4) and *a gente* (6.2.2.2), I continue to use the terms 'syntactic' and 'semantic' agreement to describe the general patterns of agreement. However, in my analysis throughout the rest of the chapter, I make use of more fine-grained distinctions, including features in INDEX and CONCORD, which are syntactic features (represented at functional structure), alongside semantic features.

In European Portuguese, *você* takes uniformly third person agreement: it has third person verb agreement (199), a third person possessive (199), and a third person direct object pronoun (200).[118] In terms of number, it is straightforwardly singular, with singular verb agreement and co-referring direct objects.

[118] In examples, I gloss *você* as 'you', without stating person features, due to its person mismatch properties.

The gender of the (third person) direct object pronoun used with *você* matches the gender of the addressee. In (200), therefore, the person being addressed as *você* is masculine in gender. Any adjectives referring to the second person agree in gender with the addressee. (The feminine possessive *sua* in (199) agrees in gender with the possessed noun *opinião*.) Although the original gender of the noun *mercê* was feminine, and therefore the source noun phrase *vossa mercê* 'your mercy' was feminine, in Portuguese, the grammaticalised pronoun is not specified for gender, showing semantic agreement.

(199)	Você	tem	a	sua	opinião	
	you.SG	have.3SG	DEF.FSG	3SG.POSS.FSG	opinion.FSG	
	You have your opinion (*e eu a minha.* 'and I mine.')					
						(*Corpus do Português*, Web/Dialects)
(200)	Você	estava	lá,	eu	o	vi.
	you.SG	be.3SG	there	1SG.SUBJ	3SG.M.DO	see.1SG
	You were there, I saw you. *(Addressing a man)*					
						(Jensen 1981, 59)

Grammatical features of the possessum can also be marked on the possessive: for the source noun phrase *vossa mercê*, the possessive *vossa* agrees in number and gender with the possessum *mercê* (see Table 39 below).

The 'conservative' pattern of syntactic agreement found with *você* in European Portuguese is also found in standard Brazilian Portuguese: *você* has third person agreement across all targets. In other Brazilian Portuguese varieties, *você* may have different grammatical features, sometimes characterised as 'incorrect' or 'misused' agreement (see for example Thomé-Williams 2004, 89). From here onwards I look specifically at two of the possible patterns of use, which are illustrated in the following examples from modern Brazilian Portuguese.

(201) Você estava lá, eu te vi.
you.SG be.3SG there, 1SG.SUBJ 2SG.DO see.1SG
You were there, I saw you.
(Jensen 1981, 59)

(202) Tenho certeza de que você vais gostar
have.1SG certainty of that you.SG go.2SG like
I am sure that you will like (it)
(*Corpus do Português*, Web/Dialects)

(203) (*Sou iniciante na cozinha mas*… 'I am a beginner in the kitchen, but…')
graças a você e teu blog
thanks to you.SG and 2SG.POSS.M blog
estou me virando muito bem.
be.1SG 1SG.REFL turn.PRPTCP very well.
Thanks to you and your blog, I am turning out very well.
(*Corpus do Português*, Web/Dialects)

One of these is a 'less grammaticalised' variant of Brazilian Portuguese, in which verb agreement remains third person, but other agreement targets (such as direct object pronouns) are second person (201). This can be described as 'mixed agreement'. The other is a marginal use found in some Brazilian Portuguese varieties, which I refer to as a 'more grammaticalised' variant. In this variant, shown in (202), semantic agreement takes precedence, and agreement is uniformly second person, including with second person verbs. This variant is not widespread and is socially stigmatised. Any use of second person verbs with *você* is regarded as 'incorrect', with prescriptive

explanations typically explaining that this is because *você* is a 'pronome de tratamento' (pronoun of address) (see for example V. M. do N. Duarte n.d.).[119]

Example (203) could be from either of these two varieties of Brazilian Portuguese, demonstrating a second person possessive (*teu*) used in combination with *você*.

I noted above (footnote 109, p277), that the form *você* is frequently used as an oblique pronoun in Brazilian Portuguese in contexts where a standard variety would use one of the clitic pronouns. This cannot be analysed as either semantic agreement or syntactic agreement, because neither the standard second person nor the standard third person coreferent pronouns are used. Instead, it represents a development of the grammaticalisation of *você* involving the extension of the form to object case, as well as subject case. I set this aside for the time being, and return to how it can be analysed from an LFG perspective, using an underspecified case feature (i.e. having the feature CASE not specified for any particular value), in section 6.4.1.

These facts together point towards substantial ongoing changes in the organisation of the pronominal system and the verb paradigm in Brazilian Portuguese. Even in a system undergoing apparently dramatic changes of this kind, and in a sociolinguistic situation involving considerable variation, the variation is structured, rather than random. Within the variation, we see important patterns of usage that a full model of Portuguese grammar should be able to account for.

To summarise, in terms of the person features of targets that co-occur with the grammaticalised pronoun, *você* in one of the Brazilian Portuguese varieties may take second person possessives and direct objects. Verb agreement can be second person (202) or third person (201), with variation across speakers and dialects. The form in (202), with second person agreement, is a more

[119] Interestingly, the use of *você* with second person verb forms is also found in African Portuguese: see Inverno (2011, 250) for discussion.

grammaticalised variant: the grammatical behaviour of the pronoun has come to match its semantic value (second person). In (201), the verb agreement is more conservative (less grammaticalised), matching the original third person feature of the full noun phrase.

In terms of other grammatical features (number and gender), the singular form *você* is uncomplicatedly singular in Brazilian Portuguese, as in European Portuguese. Second person object pronouns do not agree for gender, so gender is only marked on adjectives referring to the person addressed by *você*, which as would be expected agree in gender with the addressee, as in (204) and (205).

(204) Esse mundo é gigante e você é pequeno.
this world be.3SG huge and you be.3SG small.MSG
This world is huge and you are small. *(Addressing a man)*
(*Corpus do Português*, Web/Dialects)

(205) Você é pequena, mas não é invisível.
you be.3SG small.FSG but not be.3SG invisible.SG
You are small, but you are not invisible. *(Addressing a woman)*
(*Corpus do Português*, Web/Dialects)

Table 39 summarises the grammatical properties of the source nominal form *vossa mercê*. As described above, the 'semantics of possessum' and 'syntax of possessum' features (number and gender) for *vossa* match the source noun *mercê*.

Table 40 gives the properties of the modern form *você* as conventionally used in European Portuguese, presenting both semantic and syntactic information. Table 41 and Table 42 give the relevant equivalent information for *você* in less grammaticalised and more grammaticalised varieties of Brazilian Portuguese respectively, according to the diagnostic properties of these varieties discussed earlier in this section.

	vossa				*mercê*	
	Semantics of possessor	**Syntax of possessor**	**Semantics of possessum**	**Syntax of possessum**	**Semantics**	**Syntax**
Person	2	2	not specified	not specified	3	3
Number	singular	plural	singular	singular	singular	singular
Gender	(of addressee)	not specified	form does not have this feature	feminine	form does not have this feature	feminine

Table 39 Features of *vossa mercê*

	você	
	Semantics	**Syntax**
Person	2	3
Number	singular	singular
Gender	(of addressee)	matches semantics

Table 40 Features of *você* in European Portuguese

	você	
	Semantics	**Syntax**
Person	2	3 (some targets); 2 (some targets)
Number	singular	singular
Gender	(of addressee)	matches semantics

Table 41 Features of *você* in a less grammaticalised variety of Brazilian Portuguese

	você	
	Semantics	**Syntax**
Person	2	2
Number	singular	singular
Gender	(of addressee)	matches semantics

Table 42 Features of *você* in a more grammaticalised variety of Brazilian Portuguese

6.2.1.4 Grammatical features of vocês

The grammatical properties of the plural *vocês* are less varied. The form always occurs with third person plural verb forms. As for other agreement targets, such as the direct object, indirect object and reflexive clitics, there is some variation between the inherited third person plural forms, and the inherited second person plural forms, in European Portuguese.

Thomé-Williams gives an example of this in the context of 'misused agreement', which she states is generally more typical of Brazilian Portuguese usage of *você* (SG).

(206) Vocês trouxeram o vosso livro?
VOCÊS bring.PRET.3PL the.MSG 2PL.POSS.MSG book.MSG
'Did you (pl.) bring your book?'

(Thomé-Williams 2004, 90)

This could be considered a more grammaticalised use of *vocês*, since it is closer to semantic agreement. The fact that this pattern is only found in European Portuguese is intriguing, since generally more grammaticalised forms are typical of Brazilian Portuguese. One possible explanation is that the forms paradigmatically associated with *vós* were never widely used in Brazilian Portuguese because the inherited plural form was not in regular use by the relevant speaker groups at the time of the colonisation of Brazil. On the other hand, *vós* and its full paradigm is still used (albeit marginally) in some areas of mainland Portugal, and so speakers who have access to these forms could transfer their use over to the grammaticalised second person plural form *vocês*.

This analysis assumes that *vocês* was initially used with the third person plural forms, with the second person plural forms being a result of further grammaticalisation. However, Lara Bermejo suggests that the direction of change is the other way round: in areas of Portugal that retain *vós*, we should expect to see second person plural forms, and later developments would introduce the third person plural. Considering the forms used for multiple addressees as a whole (not distinguishing between *vós* and *vocês* as subject), he provides evidence from areas of Portugal that still retain *vós* but where *vocês* is gaining ground that the reflexive tends most towards the third person plural form, followed by the direct object, then the indirect object (Lara Bermejo 2015, 328).[120]

Putting aside the chronological question, it is clear that we are dealing with different sets of grammatical features for two variants of *vocês*. In the variant which retains more syntactic agreement, we see third person plural agreement on all targets. In the variant with mixed agreement, we see third person plural agreement on verbs, and second person plural agreement on other agreement targets (such as possessives and clitic pronouns). The general patterns align with what is expected from the combined agreement/predicate hierarchy given by Corbett (discussed in section 6.1.2 above): when change takes place, verbs retain syntactic agreement while pronouns move towards semantic agreement.

The grammatical features of these two variants of *vocês* are summarised below.

[120] Lara Bermejo also notes that some European Portuguese speakers use *vocês* as a tonic object pronoun or oblique form (2015, 328). This implies that *vocês* can have an underspecified case feature (see section 6.4.2).

	vocês	
	Semantics	**Syntax**
Person	2	3
Number	plural	plural
Gender	(of addressee)	matches semantics

Table 43 Features of *vocês* in a variety with more syntactic agreement

	vocês	
	Semantics	**Syntax**
Person	2	3 (some targets), 2 (some targets)
Number	plural	plural
Gender	(of addressee)	matches semantics

Table 44 Features of *vocês* in a variety with mixed agreement

Having seen the properties of the second person grammaticalised second person pronouns *você* and *vocês* in different varieties of Portuguese, as well as the general properties of the forms, I now move on to provide the equivalent information about the first person plural form *a gente*.

6.2.2 Data: first person plural a gente

In this section I set out the crucial facts about the grammaticalised first person plural pronoun *a gente*, from the full noun phrase of the same form meaning 'the people'. The situation of the first person plural forms is perhaps simpler than the complex range of possibilities we have just seen for second person forms. Nevertheless, the form *a gente* shows many interesting features throughout its historical development, which makes it a valuable additional case study in our analysis of mismatches associated with grammaticalised pronouns in Portuguese.

The form *a gente* has two possible functions as a pronoun: it can be a generic pronoun,[121] referring to 'people in general' like the English pronoun 'one' (207), or it can be a personal pronoun, referring to 'the speaker and one or more individuals' (208). The generic pronoun function developed first, from the use of *gente* 'people' as an indefinite or generic noun, and from the generic pronoun function, further grammaticalisation took place to produce the personal pronoun function (C. R. dos S. Lopes 2003, 13). In the examples I gloss *a gente* as A GENTE to avoid presupposing a particular analysis.

(207) O ambiente de albergue é ruim
the environment of hostel be.3SG bad
não é um ambiente para criança.
NEG be.3SG a environment for child
Também é ruim porque eles colocam
also be.3SG bad because 3PL.SUBJ put.3PL
a gente lá e esquecem de a gente.
A GENTE there and forget.3PL of A GENTE
'The hostel environment is bad, it's not an environment for children. Also, it's bad because they put people (*generic*) there and forget them'

(*Corpus do Português*, Web/Dialects)

[121] This use is sometimes referred to as an 'indefinite pronoun'. To avoid confusion with the term 'indefinite' as the opposite of 'definite' in noun phrase definiteness, I use the term 'generic' here.

(208) A gente brincava muito ali naquele pedaço
A GENTE play.IPFV.3SG much there in.that piece
'We used to play a lot around there'

(*Corpus do Português*, Web/Dialects)

The two distinct pronominal functions are not always easily distinguished. This in itself is not problematic in the identification of the form *a gente* as a pronoun, though, because the two functions are commonly shared between a single form. Even within Portuguese, the inherited first person plural form *nós* has both functions, as noted by Lopes (2003, 66). Although Lopes does not provide examples of this, the example above with *a gente* as an indefinite pronoun could be adapted to give an acceptable example:

(209) Também é ruim porque eles nos colocam
also be.3SG bad because 3PL.SUBJ 1PL.DO put.3PL
lá e esquecem de nós
there and forget.3PL of 1PL.OBL
'The hostel environment is bad, it's not an environment for children. Also, it's bad because they put us (*generic use: 'people like us'*) there and forget us'

As an example of the ambiguity of the generic and personal pronoun uses of *a gente*, Lopes gives the following example:[122]

[122] My gloss and translation.

(210)	quando	a gente	revê	isso,
	when	A GENTE	see.again.3SG	that
	a gente	relembra	essas	coisas
	A GENTE	re-remembers	these	things

'when we/one sees that again, we/one remembers these things again'

(C. R. dos S. Lopes 2003, 66)

This can be read either as referring to people in general, or to a particular group of people including the speaker. These two functions (as generic pronoun and personal pronoun) are both synchronically acceptable as uses of *a gente*.

6.2.2.1 Historical development of a gente

Historically, the generic sense of *a gente* is similar to an earlier generic pronoun grammaticalised from a noun, namely *homem* 'man', from the Latin *hominem* 'man' (accusative singular).[123] This sense of *homem* fell out of use in Portuguese in the sixteenth century (C. R. dos S. Lopes 2003, 46), and the form returned to being a simple noun. I refer the reader to Lopes (2003, 13) for further discussion.

The grammaticalisation of a noun phrase to a pronoun has various stages, as we saw above with the emergence of *você*. For *a gente*, Lopes identifies one of the earliest signs of the grammaticalisation

[123] Other Romance varieties have generic pronouns which are reflexes of the nominative singular form *homo* of this lexeme, including French *on* and Occitan *on, om* (Smith 2010, 285). The nouns meaning 'man' in these varieties (French *homme* and Occitan *ome*) are reflexes of the accusative singular Latin form *hominem*.

of the form as being the loss of the plural of the noun *gente* 'people'. In Old Portuguese, the plural form *gentes* 'peoples' occurred:

(211)	as	armas	e	caualos	que	tem
	the	weapons	and	horses	which	have
	as	gentes	dos	conqelhos		
	the.PL	people.PL	of.the	councils		

'…the weapons and horses which the peoples of the councils have…'

(C. R. dos S. Lopes 2003, 36)

The plural form was subsequently lost at around the same time that *homem* stopped acting as a generic pronoun, suggesting that the incipient function of *a gente* as a generic pronoun led to a 'blocking' of the grammaticalisation of *homem*, which went back to being a normal noun phrase with both singular and plural forms (C. R. dos S. Lopes 2003, 74).

6.2.2.2 Grammatical features of a gente

As *a gente* became more commonly used in this function as a generic pronoun, it also began to undergo further changes in its grammatical features. According to Lopes (C. R. dos S. Lopes 2003, 79), at an earlier stage of the form's grammaticalisation (the nineteenth century and earlier), it had only feminine singular agreement. An example from the nineteenth century is given here, which Lopes claims from the context clearly includes the speaker as one of the members of *a gente*, and can therefore be analysed as a pronominal use:

(212)	neste	mundo	a gente	vive	enganada
	in.this	world	A GENTE	live.3SG	deceived.FSG

'in this world, we are deceived'

(C. R. dos S. Lopes 2003, 78)

More recently, as grammaticalisation has advanced further, the form has come to act more like a standard first person plural pronoun, and a greater variation in agreement properties is found: both masculine and feminine agreement, and both singular and plural agreement, are found on different agreement targets.

Various patterns are found, with different features found on different agreement targets across varieties of Portuguese, as we might expect having seen the different properties of *você* above. Generalising across the data, more conservative varieties of Portuguese (that is, European Portuguese) retain more syntactic agreement, while more innovative varieties (some Brazilian Portuguese varieties) show more semantic agreement. With *a gente*, we sometimes find syntactic agreement: this is third person singular, reflecting the underlying features of the source noun phrase. We also find semantic agreement: this is first person, plural, and underspecified for gender, because *a gente* can refer to either the generic 'people in general', mixed gender groups or single gender groups. In general, a form that is underspecified for a particular feature form is compatible with different values for that feature, which it could otherwise have a distinct value for. In this example, the first person plural grammaticalised pronoun *a gente* can be interpreted as masculine, feminine or neither, which could be compared with the third person singular pronoun *ela* which can only be interpreted as feminine.[124]

Syntactic agreement on the verb (third person singular) is widespread, and semantic agreement is more likely to be found on adjectives, as well as coreferent possessives, reflexives and object pronouns. Again, this is in line with the predictions of the agreement/predicate hierarchy (section 6.1.2).

[124] See Kaplan (2018) for detailed discussion of underspecified features and their relationship to unspecified, indeterminate, default and unmarked features.

We therefore frequently find mixed patterns of agreement, where speakers use third person singular verbs but first person plural possessives, as in (213) from the *Corpus do Português*:

(213)	A gente	cria	nosso	próprio	universo
	A GENTE	create.3SG	1PL.POSS	own.MSG	universe.MSG
	com	o	poder	da	mente
	with	the.MSG	power.MSG	of.the.FSG	mind.FSG

'We create our own universe with the power of the mind.'

(*Corpus do Português*, Web/Dialects)

In addition, speakers may also use the whole pronominal form *a gente* as a (tonic) object and oblique form. This does not fit easily into the frame of syntactic or semantic agreement, but clearly represents a more grammaticalised form of *a gente*, which is underspecified for case. Raposo (2013, 901) states that this use is considered stigmatised by 'traditional' speakers. We have already seen one example of this use from the *Corpus do Português* in (207). Raposo also gives the following example:

(214)	A gente	gostava	que	o	professor	
	A GENTE	like.COND.3SG	that	the	professor	
	desse	uma	boa	nota	para	a gente
	give.IMPSBJV.3SG	a	good	mark	to	A GENTE

'We would like the professor to give us a good mark.'

(Raposo 2013, 901)

This can be contrasted with a more standard use, where the first person plural forms are used with *a gente*:

(215) A gente gostava que o professor
A GENTE like.COND.3SG that the professor
nos desse uma boa nota
1.PL.INDOBJ give.IPFVSUBJ.3SG a good mark
'We would like the professor to give us a good mark.'

(Raposo 2013, 901)

In terms of differences between Brazilian and European Portuguese, there is substantially less variation for *a gente* than for *você*. However, there are some differences in the frequency of use of *a gente* and the inherited form *nós*. As we might expect, given what we have seen in previous chapters, European Portuguese has a greater retention of the inherited form *nós*, while Brazilian Portuguese has an increased use of the grammaticalised form *a gente* (C. R. dos S. Lopes 2003, 123–35), particularly among younger speakers.

As for the grammatical feature properties of *a gente* across the two varieties, Marcotulio et al. (2013) argue that any differences seen in grammatical features and agreement of *a gente* between Brazilian and European Portuguese simply reflect the fact that European Portuguese is more conservative in its use of the grammaticalised pronoun *a gente* than Brazilian Portuguese. European Portuguese speakers are more likely to use *a gente* with third person and singular agreement, as are educated Brazilian Portuguese speakers, while less educated speakers are more likely to use first person plural forms (C. R. dos S. Lopes 2003, 13–14). In her corpus of Portuguese from the 1980s and 1990s, Lopes finds no examples of *a gente* with first person plural verbs among educated speakers, although it does occur with first person plural possessives and coreferent pronouns:

(216) E sai no nosso diploma que a gente
and go.out.3SG in.the 1PL.POSS.MSG diploma that A GENTE
tem condições de assinar uma planta
have.3SG conditions of sign a plan
'And it says on our diploma that we are able to sign off on a plan'

(C. R. dos S. Lopes 2003, 135)

Lopes also notes that in recent Brazilian Portuguese, it has become more common for speakers to associate *a gente* with first person plural forms, often in parallel structures rather than in the same sentence. She gives this example:

(217) A gente não se sente seguro nas
A GENTE NEG REFL.3SG feel.3SG safe.3SG in.the
ruas. Mesmo sabendo que refrescaria um
streets even knowing that refresh.COND.3SG a
pouco torcemos para não chover
little hope.1PL for NEG rain.INF
'We don't feel safe in the streets. Even knowing that it would be quite refreshing, we hope that it will not rain.'

(C. R. dos S. Lopes 2003, 137)

Examples of *a gente* with the first person plural verb form can be found, as in this example from the *Corpus do Português*:

(218)	essas	11	hrs	a gente	vamos	ter	que
	those	11	hours	A GENTE	go.PRES.1PL	have.INF	that
	trabalhar		como	loucos			
	work		like	mad.MPL			

'We're going to have to work like mad for those 11 hours'

(*Corpus do Português*, Web/Dialects)

Raposo (2013, 900) notes that this variant is 'stigmatised in the standard language' [my translation]. There is ample evidence online of prescriptive linguistic condemnation of this combination of *a gente* with first person plural verbs, such as in the online dictionary Infopedia (Porto Editora 2003a), further demonstrating that it is in current use, at least among some speakers.[125] We can therefore recognise this as a 'fully grammaticalised' form of *a gente*.

As we did for *você* above, Table 45-Table 47 present a summary of the grammatical features of *a gente* in various stages of grammaticalisation.

---

[125] The Infopedia page reads: *A palavra* gente *designa «um conjunto de pessoas», mas é um nome singular. Por isso, o verbo deve estar na terceira pessoa do singular:* A gente vai sair logo à noite? Translation: The word *gente* means 'a group of people', but it is a singular noun. Therefore, the verb must be in the third person singular:

A gente	vai	sair	logo	à	noite?
A GENTE	go.PRES.3SG	go.out.INF	soon	at.the.FSG	night.FSG

'Are we going to go out tonight?'

	a gente	
	Semantics	**Syntax**
Person	1	3
Number	plural	singular
Gender	(of addressee)	feminine

Table 45 Features of *a gente*: original noun phrase, with syntactic agreement

	a gente	
	Semantics	**Syntax**
Person	1	3 (some targets); 1 (some targets)
Number	plural	singular (some targets); plural (some targets)
Gender	(of addressee)	matches semantics

Table 46 Features of *a gente*: less grammaticalised, with mixed agreement

	a gente	
	Semantics	**Syntax**
Person	1	1
Number	plural	plural
Gender	(of addressee)	matches semantics

Table 47 Features of *a gente*: more grammaticalised, with semantic agreement

6.2.3 Interim summary

As we have seen throughout this section, the general trajectories of the second person and first person plural pronouns in Portuguese involve decreased use of the inherited forms (*tu* 2SG, *vós* 2PL and *nós* 1PL) and increased use of grammaticalised forms, although with considerable variation across Portuguese varieties. These grammaticalised forms have their roots in noun phrases: in the case of the second person, the nominal expressions were used to address a superior or distant

addressee with politeness. Importantly, there is a wide range of strategies of direct address in Portuguese, including many strategies that rely on third person address used with a nominal form (as discussed further in section 6.3.1.2). However, *você* is by far the most grammaticalised of second person address forms with origins as noun phrases. In the first person plural, the form *a gente* from the source noun phase meaning 'the people' is an extension of its use as a generic pronoun, as it came to designate an in-group.

Alongside an increased use of these grammaticalised forms, their grammatical features develop new complexities, most clearly in the case of *você* where we see considerable variation across Brazilian Portuguese varieties in the marking of second person versus third person singular on a variety of agreement targets (including verbs, possessives and object pronouns).

An effective formal analysis of Portuguese needs to be able to account for the different patterns that are seen in different varieties, including differences in the precise nature of the mismatch of grammatical features and of semantics versus syntax. We also need to capture the change in reference associated with the transition from noun phrase to pronoun. However, the change in meaning is not accompanied by instant, clear-cut changes at all levels of grammatical analysis.

We have seen here that a variety of changes and properties need to be included within the analysis, including category membership, mismatches between syntax and semantics, and mismatches between grammatical features at different levels of the syntax. The purpose of section 6.3 is to set out the ways in which these phenomena can be accounted for in the formal LFG-PFM analysis, setting the scene for an analysis of the Portuguese case studies in section 6.4.

6.3 A formal LFG-PFM account of grammaticalised pronouns

This section presents the main elements of my formal analysis of the grammaticalised pronouns in Portuguese. I demonstrate how the LFG-PFM approach can capture the distinct processes at work in the historical trajectories of these forms as they grammaticalise. The aim of this section is

to present some different elements of the formal analysis, setting the scene for section 6.4, which applies these elements to the Portuguese case studies of *você*, *vocês* and *a gente*.

This section considers several aspects of the grammaticalisation process separately. In section 6.3.1, I look at the representation of changes in lexical category in both LFG and PFM, since category change can be perceived as relevant in the trajectory from source noun phrase to pronoun. Section 6.3.2 considers the pronouns from a PFM perspective, looking at the properties of pronouns' inflectional paradigms, and considering how these are represented within rule blocks for nominal elements more generally. Section 6.3.3 turns to the representation of mismatches between semantics and syntax. This type of mismatch arises as part of the grammaticalisation process: for the forms studied here, there is a stage where they have the semantic properties of a pronoun, but retain (some of) the syntactic properties of a noun phrase. This type of mismatch can be represented in LFG by the use of elaborated semantic structures, alongside functional structures.

Finally, in section 6.3.4, I consider the representation of grammatical features in LFG, with a particular focus on the representation of feature mismatches using INDEX and CONCORD, the feature bundles introduced in section 6.1.3 above. This section also touches on the interaction between LFG features and the PFM conception of features: since PFM deals with purely morphological features, it is possible for there to be mismatches between the two, as we first saw in Chapter 2.

6.3.1 Relationship between nouns and pronouns in the formal framework

We saw in Chapter 4 that the distinction between the lexical categories of verbs and adjectives may be relevant for accounting for some of the properties of past participles. We have also seen that lexical category plays a role in both LFG, being listed in the lexemic entry and determining c-structure properties, and in PFM, where morphological category determines which inflectional rules will apply. More broadly, category change is often part of the process of grammaticalisation:

a lexical item may grammaticalise from a lexical category to a more functional category, as has been considered in some detail in previous LFG work (see section 6.3.1.2). On the other hand, category change has rarely if ever been considered in PFM, as it is chiefly a synchronic theory.

In Portuguese, Raposo (2013, 889–92) argues that pronouns are syntactically and semantically equivalent to noun phrases, not nouns. Syntactically speaking, pronouns replace the whole noun phrase including determiners and modifiers, and semantically, they 'inherently combine the semantic value of a specifier (determiner or quantifier) with a descriptive meaning with nominal properties' (Raposo 2013, 890–91).[126] Raposo states that, for example, third person pronouns can be considered to 'incorporate the meaning of the definite article', since they pick out a particular referent. This aligns clearly with the grammaticalisation process by which, for example, the noun phrase *a gente* 'the people' becomes the pronoun *a gente*. There is no particular semantic or syntactic change in category, but there is a change in reference. The LFG representation of this process, using the PRED feature, is discussed in section 6.3.1.3.

In section 6.3.1.1, I discuss the status of the grammaticalised forms as nouns or pronouns in more detail. I then consider the grammatical status of the 'nominal' forms of address in 6.3.1.2. I make some suggestions for how the relationship between nouns and pronouns should be considered from a PFM perspective in section 6.3.2.

[126] Lopes (2003, 23–28) discusses differences between 'substantive nouns' and personal pronouns in Portuguese, although the distinctions that she makes, generally involving the relationship with determiners and modifiers, seem to be accounted for by considering pronouns to be equivalent to full noun phrases, rather than nouns.

6.3.1.1 Status of grammaticalised forms: noun phrases or pronouns?

There is no consensus on the question of a category distinction between pronouns and nouns. As we have seen above, lexical category questions involve a number of distinct factors, which can result in inconsistent conclusions if different factors are conflated. Sugamoto (1989a; 1989b) has suggested a continuum between nouns and pronouns, with elements usually identified as 'pronouns' in some languages being categorically identical to nouns. In some theoretical approaches, different types of pronouns are analysed as representing different syntactic categories, with some being equivalent to nouns (Déchaine and Wiltschko 2002; 2003). In the LFG approach which I follow, personal pronouns and relative pronouns are generally treated as having the syntactic category N.

Taking an empirical view, there are clear similarities between pronouns and nominal forms, but also important differences. Semantically, one difference is that pronouns have anaphoric reference, while nouns typically have a direct reference. Heine and Song (2011, 589) argue that personal pronouns (cross-linguistically, not just in Portuguese) are a distinct category from nouns, due to their particular referential properties and restricted distributional properties (compared with nouns). The semantic properties of personal pronouns clearly need a different account from that used for nouns: see footnote 143 in section 6.4.4 below.

Syntactic criteria have also been proposed to identify pronouns. Three tests that have been suggested are 'adjectival modification', co-occurrence with numerals, and the 'we men' test (M. Taylor 2009). Pronouns typically do not take adjectival modification (with the exception of special cases like 'poor me'), but they do co-occur with numerals (e.g. 'we three', 'you two'). In the final test, non-third person pronouns are argued to be able to occur as the head of a DP, as in the English example 'we men'.

The forms *você* and *vocês* straightforwardly meet all these criteria, and can be considered as grammaticalised pronouns that are well integrated into the Portuguese grammar. On the other

hand, the form *a gente*, which is generally considered also to be a pronoun (see for example Raposo 2013, 899), has recently been scrutinised more closely by M. Taylor (2009), who provides evidence of contexts where *nós* cannot be straightforwardly replaced by *a gente*, referring to the 'we men' context (219), (220) and co-occurrence with numerals (221), (222).

(219) Nós brasileiros não gostamos de vocês portugueses.
we Brazilians not like.1PL of you Portuguese
'We Brazilians don't like you Portuguese.'

(M. Taylor 2009, 7)

(220) *A gente brasileiros não gosta de vocês portugueses.
A GENTE Brazilians not like.3SG of you Portuguese
'We Brazilians don't like you Portuguese.'

(M. Taylor 2009, 7)

(221) Nós três fomos à igreja.
we three went.1PL to.the church
'We three went to church.'

(M. Taylor 2009, 6)

(222) *A gente três fomos/foi à igreja.
A GENTE three went.1PL/3SG to.the church
'We three went to church.'

(M. Taylor 2009, 6)

M. Taylor argues that there are actually two forms of *a gente*, neither of which is a straightforward pronoun. *A gente* is 'a more complex syntactic object that contains the silent 1 person plural pronoun NÓS'. To complicate matters further, he argues that there are in fact two co-existing constructions for *a gente*: a standard one (which takes third person singular verb agreement) and a non-standard one (with either third person singular or first person plural verb agreement). He analyses the standard form as an 'imposter', under the definition of Collins and Postal (2012), and the non-standard form as a 'pluringular' (Dikken 2001), or more formally, a complex DP which includes a 'non-standard appositive construction':

(223) $[_{DP}\,\text{NÓS}_{pl}\,[_{DP}$ a gente [sg]]]

Costa and Pereira (2013) argue that the standard analysis of *a gente* as a pronoun should be retained, although they accept some of Taylor's analysis of *a gente*. They state that 'the only difference [between their analysis and Taylor's] is on the status of the word *gente*, which we treat as part of the D head, and not as an independent noun, that is, the expression *a gente* is a grammaticalized form without internal structure'. They propose a Minimalist phase-based explanation to account for the differences between *nós* and *a gente* in behaviour with numerals and the 'we men' context. The singular number feature of *a gente* conflicts with the plural numeral or plural noun in the same local domain. This is a result of the 'pluringular' (mismatched number) properties of *a gente*. The effects can be eliminated by inserting a determiner to act as a 'phase boundary', creating an apposition, which allows semantic agreement to occur:

(224) A gente os brasileiros não gosta de vocês portugueses.
A GENTE the Brazilians not like.3SG of you Portuguese
'We Brazilians don't like you Portuguese.'

(225) A gente os três fomos/foi à igreja.
A GENTE the three went.1PL/3SG to.the church
'We three went to church.'

(Adapted from M. Taylor (2009), in line with Costa and Pereira (2013))

According to this analysis, the controversy over the status of *a gente* as a pronoun is unnecessary; the only examples in which *a gente* appears not to act like a pronoun are accounted for by constraints on agreement which become apparent due to mismatches occurring in a local domain.

Taking into account these possible analyses, it remains clear that *a gente* (like the more straightforward pronoun *você*) can be considered as an N in terms of syntactic category: from an LFG perspective, it has nominal properties.[127] Stepping back, the purpose of this chapter is to consider the mismatch behaviour of the grammaticalised pronouns, particularly for the features of person and number. My LFG analysis therefore considers some of the same issues considered in Taylor's and Costa and Pereira's analyses, but from a different perspective, separating grammatical features from lexical category, given the modular architecture assumed here.

I next consider briefly the status of the 'nominal' forms of address.

[127] At c-structure, both DP and NP analyses are available (Dalrymple, Lowe, and Mycock 2019, 100–103).

6.3.1.2 Status of 'nominal' forms of address

As mentioned in section 6.2.1.1, Portuguese has many possible 'nominal' forms of address, also known as 'descriptive pronouns'. According to Amaral (2019) these are noun phrases, consisting of a definite article followed by 'a proper name, a kinship term or a professional title', interpreted as direct address forms. Amaral (2019, 59) notes their unusual combination of properties:

1. They refer to the addressee in a specific conversation, i.e. they have a deictic component;
2. They introduce a property of the addressee, e.g. having a certain name or profession, i.e. they have a property component;
3. They convey an attitude of the speaker towards the addressee, i.e. they have an expressive component.

Amaral gives the following example:

(226) [Addressing Pedro/the speaker's father/a university professor]

O	Pedro/ o		Pai/	o	Professor
the.MSG	Pedro/ the.MSG		father/	the.MSG	professor
quer	um	café?			
want.PRS.3SG	a	coffee			

'Do you [lit. the Pedro/the father/the Professor] want a coffee?'

(Amaral 2019, 58)

In the same category we could also include the very common polite forms of address *o senhor/a senhora* 'the gentleman/the lady' and noun phrases like *a menina* 'the girl' which do not fit into the categories of 'proper noun, kinship term or professional title' but are used as forms of address (see for example Lara Bermejo and Guilherme 2018, 345).

Amaral (2019) provides evidence for these forms being pronouns, not nouns: they cannot be modified by an adjective (in the relevant sense, although the identical 'ordinary' noun phrases can

be modified with adjectives e.g. *o outro senhor* 'the other man', *a pequena menina* 'the small girl'), and they are a closed class, with terms like *o filósofo* 'the philosopher' not being acceptable when used in this way. Their distribution is identical to inherited personal pronouns like *tu*. For these reasons, there is scope for a syntactic analysis of these nominal address forms which is analogous to the grammaticalised pronouns considered in chapter. I do not pursue this possibility here, focusing instead on the grammaticalised forms *você(s)* and *a gente*.

In the next section, I consider how the changing reference of the forms *vossa mercê* > *você* and *a gente*, as they grammaticalise from straightforward noun phrases to pronouns, is accounted for in the LFG framework.

6.3.1.3 Nouns, pronouns and the PRED *feature in LFG*

The process of grammaticalisation, by which straightforwardly lexical items come to have a more functional value, often brings about category change. In LFG, grammaticalisation has been considered as part of wider work on extending the theory to cover diachronic change and processes. Within this work, 'the loss of a PRED feature is a key characteristic of grammaticalization' (Dalrymple, Lowe, and Mycock 2019, 730), marking out the transition from lexical item to functional item. For example, Vincent (1999) demonstrates that the loss of the PRED feature is crucial to the process by which prepositions can come to have the function previously fulfilled by grammatical case marking, and come to represent purely grammatical relations. An example is the preposition *de* which, from Latin to Romance, undergoes grammaticalisation from a semantic value ('down, about'), to marking nominal dependents, a function previously carried out by the genitive case (Vincent 1999, 1143). The value of the PRED feature is thus closely linked to grammaticalisation.

As stated above, pronouns come under the category of nouns in LFG, and a change in lexical category is not required. However, some change in the PRED value of the forms is needed to account for the grammaticalisation of noun phrases into pronouns in Portuguese. The PRED

feature is not lost: a pronoun with no PRED value would be semantically empty, which is not appropriate for personal pronouns with anaphoric reference.[128] The PRED value of a pronoun in LFG is given as 'PRO', which represents the anaphoric semantic value of the form. The value PRO requires an antecedent which specifies the identity of the referent, meaning the value of the pronoun is dependent on another element. This can be viewed as semantically bleached to some degree, compared with the fully explicit PRED value associated with a full noun phrase (with independent meaning). The following partial f-structures demonstrate the difference between the PRED value of a full noun phrase, compared with that of a pronoun:

(227) F-structure for 'Jane'

$$[\text{PRED} \quad \text{'JANE'}]$$

(228) F-structure for 'she'

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{PERS} \end{bmatrix}$$

The PRED value 'PRO' is something of a halfway house between a fully lexical meaning and a fully grammatical one. This is appropriate, since pronouns are neither as fully lexical as nouns nor as fully grammatical as affixes, and require an antecedent, while being syntactically independent. We can conceptualise this change as below:

[128] A pronoun can lose its PRED value 'PRO' and become an agreement marker with no particular pronominal value. See Coppock and Wechsler (2010) for a discussion of this process, as mentioned in section 3.2.2. Berman (1999; 2003) also proposes that verbal morphology in German cannot introduce a PRED value (i.e. a semantic feature), but can introduce an expletive subject that has only agreement features.

(229)

Noun phrase	>	pronoun	>	affix
Lexical PRED	>	PRED 'PRO'	>	no PRED

The change in category from noun to pronoun in LFG is therefore reflected in the loss of the lexical value of the PRED feature, to be replaced by the value PRO. This is demonstrated for the grammaticalised pronouns under discussion in this chapter in section 6.4, where the PRED values make up part of the f-structure for each form under discussion (source noun phrase and grammaticalised form).

Finally, we should note that the grammaticalisation process involves many other changes alongside the change in PRED value, such as in phonetic form and in other grammatical features. Each of these changes is fundamentally independent, and the change in PRED value does not necessarily occur simultaneously with the other changes. The question of whether the process of grammaticalisation is identified by a change of PRED value or by one of the other changes is perhaps definitional, although empirical data could shed light on whether in practice changes in reference precede other grammatical or phonetic changes.

Having seen how this aspect of the change from noun to pronoun is represented in LFG, I now move on to consider how it is represented in the PFM part of my analysis.

6.3.2 PFM analysis of nouns and pronouns

As mentioned in Chapter 4, morphological category is relevant within PFM because PFM rules typically specify which the category of items they apply to. For example, the rules for verbs (given in Appendix III and Appendix IV) specify that they apply to verbs in Portuguese. One possible approach to the PFM treatment of pronouns would be to include pronouns as nouns, and to have rule blocks that account for the inflection of pronouns alongside nouns. Alternatively, we could specify different blocks of rules for nouns and pronouns.

For the purposes of this work, I propose to carry over the LFG treatment of pronouns as members of the lexical category N into PFM, and thus use the same rule blocks for nouns as for pronouns. However, we must deal with two complications: first, that pronouns have highly idiosyncratic inflectional patterns, and second, that the features they inflect for are different compared with nouns. This means that in effect, there will be different rules applying to pronouns than to nouns. They can be included in the same rule blocks, but due to the normal principle of selecting the narrowest applicable rule, only the rules that are relevant to pronouns (and not those that are only relevant to nouns) will apply. The types of rules to be proposed are discussed in more depth below.

To justify my decision in including pronouns in rule blocks for nominal inflection, it is worth first considering what type of PFM rules apply to canonical nouns in Portuguese. Portuguese nouns have a small inflectional paradigm, with only two possible values (singular or plural) for the number feature. Nouns do not inflect for case, or for gender: semantically minimal pairs with different gender specifications exist (e.g. *menino* 'boy', *menina* 'girl') but are generally treated as related by derivation, not inflection.[129] This means that the only rule that applies regularly to nouns is pluralisation.[130]

---

[129] An analysis using Spencer's (2013) *Lexical Relatedness* theory might shed more light on the exact relationship between nouns of this kind. I leave this for future work.

[130] Many Portuguese nouns also have diminutive forms: for a concise description, see Villalva and Gonçalves (2016). Diminutives are an example of evaluative morphology, which might be treated as inflectional, derivational, or an intermediate form of lexical relatedness. See Grandi (2015) for an overview. Evaluative morphology is included as an intermediate form of lexical relatedness in Spencer's (2013) account, summarised in Table 7 above. A closer examination of diminutives in

Not all Portuguese nouns have a plural form: like in English, mass nouns have no plural. In general, however, canonical nouns inflect for number. By contrast, the inflectional paradigm of pronouns in Portuguese should not include number, but should include the different case features of the pronoun. Beyond this, they should be separate forms.[131] For example, we should not inflect a single shared root to get both the singular form *você* and the plural form *vocês*. Although there is clearly a relationship between the singular and plural forms, listing each person and number combination pronoun individually in the lexicon provides a simpler way of negotiating the many choices involved in selecting the appropriate pronominal form. For example, neither the inherited first person plural pronoun *nós* nor the grammaticalised first person plural pronoun *a gente* can legitimately be described as the plural of the first person singular pronoun *eu*. They have no morphological similarities, and their reference and pragmatic properties are also different. On the other hand, the clitic direct object form *nos* can straightforwardly be viewed as an alternative morphological form of the subject form *nós*, with closely related morphology, meaning and sociolinguistic properties. Of course, there are substantial morphological idiosyncrasies in the case inflection of the different pronouns, meaning that the PFM rules must involve a considerable amount of lexical specification.

The alternative 'extreme' possibility would be having one set of paradigm rules to generate all pronominal forms with different person and number features, as well as case. This seems

---

the context of the lexical relatedness of nominal forms in Portuguese would be worthwhile but is beyond the scope of this thesis.

[131] Applying Spencer's (2013) analysis, this would mean that each of the person/number specifications for the pronouns has a separate LI (Lexemic Index), while the different case forms of a single pronoun share an LI. This seems a reasonable solution.

undesirable, not least due to the complexities of selecting from competing forms (such as *tu* versus *você*) which involves pragmatic (not syntactic or morphological) factors.

Broadly, we might expect the morphology of the grammaticalised pronouns to behave similarly to the inherited pronouns. However, even within the inherited pronouns, there is substantial variation in the number of different forms associated with the different pronouns. For example, the third person has more morphosyntactic distinctions than the first and second person, including separate forms for the direct object, indirect object and reflexive or reciprocal pronoun, as we saw in Table 34 above, repeated here for convenience as Table 48 (adapted from Raposo et al. 2013, 1:902).

		Stress-bearing (tonic) forms			Clitics (atonic forms)		
Person and number		**Subject**	**Oblique**	**Possessive**	**Direct object**	**Indirect object**	**Reflexive or reciprocal**
1SG (inherited)		*eu*	*mim*	*meu, minha, meus, minhas*	*me*		
2SG (inherited)		*tu*	*ti*	*teu, tua, teus, tuas*	*te*		
2SG (grammaticalised)		***você***	***você, si***	***seu, sua, seus, suas***	***o***	***lhe***	***se***
					a		
3SG (inherited)	M	*ele*	*ele, si*	*seus, suas, seus, suas*	*o*	*lhe*	*se*
	F	*ela*	*ela, si*	*seus, suas, seus, suas*	*a*		
1PL (inherited)		*nós*	*nós*	*nosso, nossa, nossos, nossas*	*nos*		
1PL (grammaticalised)		***a gente***	***nós***	***nosso, nossa, nossos, nossas***	***nos, se***		***nos, se***
2PL (inherited)		*vós*	*vós*	*vosso, vossa, vossos, vossas*	*vos*		
2PL (grammaticalised)		***vocês***	***vocês***	***vosso etc., seu etc.***	***os***	***lhes***	***se***
					as		
					vos		
3PL (inherited)	M	*eles*	*eles*	*seu, sua, seus, suas*	*os*	*lhes*	*se*
	F	*elas*	*elas*	*seu, sua, seus, suas*	*as*		

Table 48 Pronoun forms in Portuguese: full paradigms

The grammaticalised pronouns generally show less complex morphological patterns than the inherited pronouns: they have no inflectional morphology, with the same forms fulfilling all functions, and they do not have a clitic form. Although Raposo represents them as having a full set of associated possessives and clitic pronouns, these are in fact the forms associated with the standard inherited pronouns, as can be verified in Table 48.

There are some difficulties involved in treating the different forms of the pronouns as representing different cases, not least because some of the forms involved (direct object, indirect object and reflexive/reciprocal) are clitic forms. The behaviour of Portuguese clitics raises many questions. The same forms (direct object, indirect object and reflexive/reciprocal) occur in three positions, as discussed briefly in section 3.1.1, depending on morphosyntactic context. They can be positioned preverbally ('proclisis'), postverbally ('enclisis'), and between the verb stem and tense suffixes ('mesoclisis'), with differences in syntactic behaviour, such as scope, associated with the different positions. Although they are associated with identical f-structure information, they have a different role in the c-structure.

Luís (2004) presents a PFM analysis to account for this complex behaviour. She analyses the clitics as affixes that act as verbal suffixes when they occur in the post-verbal position and phrasal affixes in the pre-verbal position, to account for many syntactic differences in scope and other aspects. She also develops an explanation in LFG to account for the apparent breaches of the Lexical Integrity Principle that are needed to account for the behaviour of proclitics, which may appear not adjacent to the verb, but are treated as phrasal affixes: this solution involves proclitics not being included under any particular constituent structure node, but being notionally attached via 'phrasal affixation' which allows them to contribute information to the functional structure. Luís recognises that her solution leaves open questions for the phonological string (which relies on having a correspondence with the linear order captured in constituent structure). Later LFG work develops a distinction between the phonological string and the syntactic string (Dalrymple and Mycock 2011), which may help to resolve this tension. However, the analysis of verb + clitic combinations which are treated as a single morphological form and a discontinuous syntactic constituent is clearly not straightforward, and may be better treated at the level of the interface between morphology and syntax rather than at the level of the interface between phonology and syntax.

A further possible solution to the mismatch between the morphological form and syntactic behaviour of the clitics is offered by Luís and Otoguro (2005), who propose a labelling algorithm that mediates the interface between morphological tokens (produced by PFM) and syntactic atoms (which are inserted into the c-structure), a proposal that has similarities with the interface to morphology applied in this thesis (Dalrymple 2015; Dalrymple, Lowe, and Mycock 2019, chap. 12). Due to the complexities of the behaviour of clitics in Portuguese and the additional requirements to account for them in both the morphological component and the morphology-syntax interface, I do not include Luís and Otoguro's analysis of the clitics in my analysis of Portuguese morphology, but refer the reader to the works cited for further details.

Returning to the question of case in my LFG-PFM analysis, where pronouns do not have distinct forms for the subject and object, such as the grammaticalised forms *você*, *vocês* and *a gente*, this can be dealt with by saying they are underspecified for case at the level of f-structure (as I shall discuss in section 6.4), whereas pronouns with distinct case forms will have a case feature at f-structure, and will be subject to PFM rules that specify the form used in each case. The PFM rules involved must also only apply to the relevant pronouns, listing only the relevant form in the rule (see section 6.4.5).

It may therefore appear that PFM has little to say about the morphology of the grammaticalised pronoun forms because these forms do not have inflectional paradigms as described here (in the sense of different forms varying across different cases). However, the morphological component plays an important role in the production of complete lexical entries, identifying the correct pairings of morphosyntactic features with whole wordforms. This points towards the usefulness of a PFM conception of morphology as part of the wider LFG architecture which deals with different types of grammatical behaviour in separate modules.

In the next section, I move on to consider how LFG can capture mismatches between syntax and semantics, which are an intriguing part of the grammaticalised pronouns' behaviour.

6.3.3 Mismatches between syntax and semantics and their representation in LFG

This section considers the representation of mismatches between syntax and semantics of the grammaticalised pronoun forms. At some stage in their development, these forms have the typical semantics of a pronoun (that is, they are interpreted as meaning 2SG, 2PL or 1PL) but they have syntactic qualities more associated with noun phrases (such as being grammatically third person). As the forms grammaticalise further, this mismatch between syntax and semantics may eventually be eliminated, although this is by no means assured. Linguistic changes of any kind, including grammaticalisation as well as sound changes and changes at other levels of structure, are not teleological, and may stop or become 'frozen' at any point, leading to the existence of stable synchronic mismatches. The extent to which grammaticalisation-type changes may reverse has been hotly debated in historical linguistics under the name of the 'unidirectionality hypothesis' (see for example Traugott 2001; Norde 2009; 2010; Börjars and Vincent 2011). Taking a particular position on this issue is not necessary to be able to state that the start of a grammaticalisation change in one direction does not mean that that change will continue to a particular end.

The Portuguese pronouns are an interesting case in point. We see a variety of possible outcomes for pronouns which have undergone grammaticalisation to some extent, particularly in terms of their grammatical features as reflected on different agreement targets. It is not unusual for there to be a 'frozen' mismatch between the semantic properties of the pronoun and the agreement patterns it shows morphosyntactically. As we have seen, this type of mismatch is seen in *você* in European Portuguese and some Brazilian Portuguese varieties, as well as in *a gente* for many speakers. These possible outcomes are represented in Figure 15 below, using *você* as a case study.

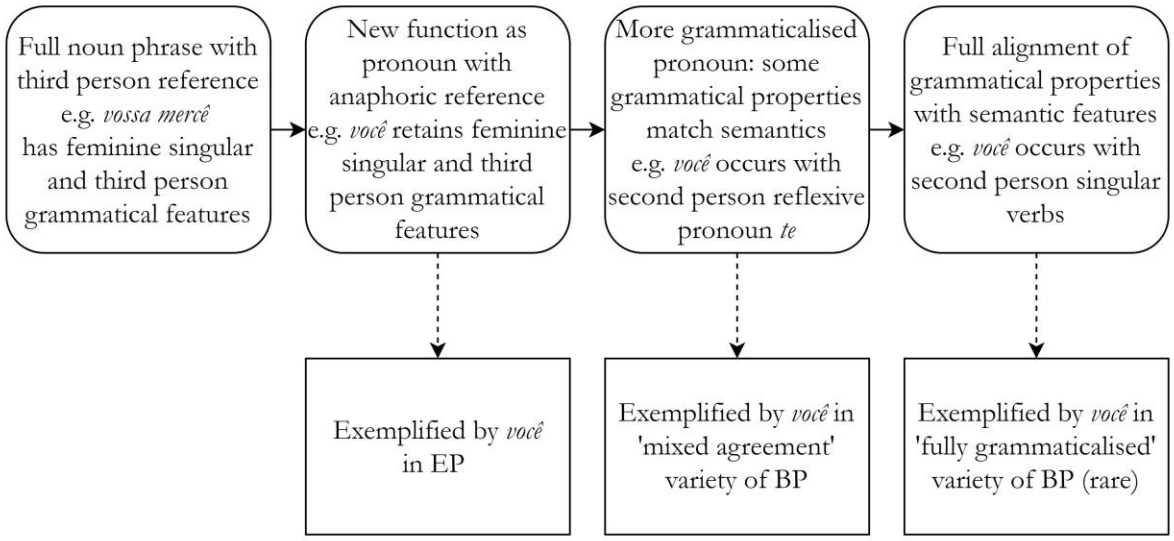


Figure 15 Possible outcomes of grammaticalisation process for pronouns

On the other hand, some marginal Brazilian Portuguese varieties show the possible extreme outcome of the mismatch being eliminated. In these varieties, *você* comes to behave grammatically like a second person singular form, including with second person agreement on verbs. This outcome is also seen in some Brazilian Portuguese varieties with the first person plural form *a gente*. Interestingly, however, this extreme grammaticalisation outcome is never seen for the second person plural (which would give us second person plural inflected verb forms with the form *vocês*). This is likely to be due to a general loss of second person plural verbal inflection in many Portuguese varieties, which probably occurred due to the widespread dominance of nominal address forms for the second person plural, which take third person plural verb agreement (see section 6.2.1.4). Apparent mismatches between syntax and semantics therefore call into question exactly what we are talking about when we use terms like 'second person plural verb agreement', if a different verb form is always used for second person plural referents. This would imply that the morphological rules should be adapted to account for the fact that (in this case) the third person plural verb is used for the second person plural as well, an instance of syncretism. This has implications for the wider morphosyntactic analysis of Portuguese, recognising that morphological changes (such as the reduction of person and number distinctions in verb paradigms) and syntactic

changes (such as the loss of inherited pronouns in favour of grammaticalised forms) have considerable interactions. The LFG architecture represents a useful perspective from which to make such analyses. A full analysis is beyond the scope of this chapter, but I suggest some directions for continuing this work in section 6.5.3.

In any case, it is clear that at some stages of the grammaticalisation process there are important mismatches between syntax and semantics. These mismatches are captured in my analysis using LFG functional structures and semantic structures, which represent attested changes from syntactic agreement with the source nominal form towards semantic agreement with the features of the referent.

To be able to capture the semantics-syntax mismatches, and the changes in syntactic features that occur through the grammaticalisation process, it is necessary to represent both the intrinsic 'semantic' and discourse-relevant pragmatic properties of the addressee, including person, number and gender features, as well as the syntactic features that the pronoun may show on different agreement targets. Partial semantic structures in the form of attribute-value matrices are therefore given that include information about PERSON and NUMBER, to capture the fact that the semantic features at this level may be different from both the INDEX and CONCORD features represented at the f-structure level of syntactic analysis.[132] This distinction can be found for example in pronouns whose reference (semantics) is clearly second person, but whose agreement patterns (which are syntactic) remain purely third person, like *você* in European Portuguese.

[132] As discussed above in section 6.1.3 and further in section 6.3.4 below, in principle there can be splits between INDEX, CONCORD and semantics so all three must be treated separately, following Wechsler and Zlatić (2000).

Including such information at the level of s-structure is an innovation within LFG. However, there is a precedent for adding more detail to these representations. Dalrymple and Nikolaeva (2011), partly following Liao (2010), state that semantic structures can contain 'a rich array of semantic and pragmatic features', including animacy, identifiability, and activation status in the discourse, and note that 'other features might be relevant in other contexts.' The representation of relevant feature information in the s-structure allows us to represent more accurately the behaviour of the different pronouns and capture how this may change over time. Such information is not typically represented in the s-structure, but has been implicitly assumed (see for example s-structures in Dalrymple, Lowe, and Mycock 2019, 312–13).[133] I also include information about animacy (following Liao 2010) and the ±HUMAN feature, to exemplify a complete representation of relevant information at s-structure.

6.3.4 Formal representation of grammatical features

The grammatical features of person, number and gender are generally represented in LFG at the level of functional structure. As we saw in section 6.1.3 above, these features are sometimes 'bundled' together in two distinct groups, INDEX and CONCORD, and a single noun may be specified for contradictory features within the two feature bundles. For example, in cases of number and gender mismatch like the Serbo-Croatian noun *deca* 'children', a noun can be specified

---

[133] Kaplan's comments about syntactic representations could be equally applied to semantic structure: 'there is a preference in LFG for smaller syntactic representations that are not cluttered with predictable features and values. […] Unremarkable values of certain features can also be systematically omitted from syntactic structures in a way that reduces uninformative clutter but still allows syntactic dependencies to be enforced' (2018, 356). Where no mismatch is involved, the features I include can be characterised as predictable or unremarkable.

for neuter plural INDEX features and feminine singular CONCORD features (Wechsler and Zlatić 2000, 816), as represented here in an LFG functional structure:

(230) Sample partial f-structure for Serbo-Croatian *deca* 'children'

$$\begin{bmatrix} \text{PRED} & \text{'children'} \\ \text{INDEX} & \begin{bmatrix} \text{GEND} & \text{N} \\ \text{NUM} & \text{PL} \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{GEND} & \text{F} \\ \text{NUM} & \text{SG} \end{bmatrix} \end{bmatrix}$$

Grammatical features may also be represented in LFG at the level of lexical and lexemic entries. Lexemic entries contain features which are shared by the whole paradigm. For example, as discussed in section 6.3.1, the English first person singular pronoun 'I' would be specified at the level of the lexemic entry to have the grammatical features NUMBER: SINGULAR and PERSON: 1. On the other hand, lexical entries contain grammatical features that are not shared by the whole paradigm of the lexeme, but only by specific wordforms. For example, the English plural noun 'dogs' would have the feature NUMBER: PLURAL specified in its lexical entry.

The mismatches addressed in this chapter are generally either between semantics and syntax, or within two different aspects of syntax. They do not involve purely morphological features. For this reason, the mismatches are captured at the level of f-structure and s-structure. These levels of structure do not interact with the morphological component of Paradigm Function Morphology; in PFM, we need only refer to morphological features. As we saw in Chapter 2, there can be mismatches between syntactic features (s-features) and morphological features (m-features), which are dealt with by the description function $\mathcal{D}$: an example is deponent verbs in Latin, which are passive in form but active in meaning. The existence of this type of mismatch justifies the distinction between the two sets of features.

The existence of four distinct sets of features to treat mismatches was first proposed by Wechsler and Zlatić (2000, 800). Of the four sets, three are relevant to the types of mismatches discussed in

this chapter: semantic features, INDEX features and CONCORD features. The four sets of features are summarised in Table 49 below, arranged from 'most semantic' to 'most morphological'.

Type of feature	**Description**
Semantic features	The real-world properties of the referent
Index features ('referential index')	Closely related to the semantics, but a syntactic feature
Concord features	Another syntactic feature, normally corresponds with morphological form
Morphological features ('declension')	Involved in determining morphological properties (wordforms)

Table 49 Four distinct sets of features to analyse mismatches

In principle, any of these features can be mismatched with their neighbours; as Wechsler and Zlatic describe (2000, 800), in general there is alignment across all four features, but a word can have a split at any point, and even, exceptionally, two splits: see p289 above for discussion of the example of Serbo-Croatian *braća* 'brothers' (Wechsler and Zlatić 2000, 814–15).

Having seen the basis of the formal analysis, in the next section I apply this frame of analysis to the Portuguese grammaticalised pronouns studied here: the second person forms, both singular and plural, and the first person plural forms.

In the examples that follow in sections 6.4.1, 6.4.2 and 6.4.3, semantic features are represented in semantic structures, as discussed in section 6.3.3, while INDEX and CONCORD features are represented at the level of f-structure. We see examples of mismatches between semantics on one hand and INDEX and CONCORD on the other (231), and between semantics and INDEX on one hand and CONCORD on the other (232).

(231) DECLENSION ↔ CONCORD ↔ INDEX | SEMANTICS

(232) DECLENSION ↔ CONCORD | INDEX ↔ SEMANTICS

A simple 'syntactic/semantic' agreement split is unable to account for both varieties with uniformly syntactic agreement across all targets, and for varieties where some targets show semantic agreement, necessitating the use of INDEX and CONCORD features. As well as the semantic and functional structures, I provide lexical entries for the grammaticalised pronouns in section 6.4.4, and sketch some PFM rules for the paradigms of pronouns in 6.4.5.

6.4 Worked formal representations

This section takes forward the previous section's account of how various formal aspects of the morphology, syntax and semantics of grammaticalised pronouns can be represented in LFG-PFM and applies it to the Portuguese case studies. I consider worked examples of several kinds for both second person pronouns (sections 6.4.1 and 6.4.2) and first person plural pronouns (section 6.4.3), giving examples of semantic structures and functional structures. In section 6.4.4 I present sample lexical and lexemic entries for Portuguese pronouns, and in section 6.4.5 I show how PFM rules could account for the inflectional morphological properties of the pronouns. The overall aim of this section is to capture changes in both grammatical features and semantic information carried by the forms as they grammaticalise from nouns to pronouns.

The formal analyses given below include LFG f-structures, making use of the INDEX and CONCORD distinction, as described in section 6.1.3, and s-structures, as outlined in section 6.3.3. I implement two innovations: at the level of f-structure, the inclusion of PERSON in CONCORD (which has previously contained only the features CASE, NUMBER and GENDER), and at the level of s-structure, the addition of information about PERSON, NUMBER and GENDER (see section 6.3.3).

The structures presented below sometimes contain features with a blank value. Since the examples presented here are for abstract lexical items given out of context, with no particular referent of the

pronoun in mind, more than one value is possible for a given feature. An example is the value of the feature GENDER: unless the gender of the second person addressee is known, it is not possible to give a value to the GENDER feature. In this case, I leave the feature value blank, following common practice in the LFG literature.[134] This represents underspecification, accounting for the fact that the value varies depending on the context for the individual instance of the pronoun.

As further background, I present here a sample f- and s-structure for a simple pronoun, in this case the English third person pronoun 'he', incorporating the innovations I propose:

(233) F-structure for 'he'

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{PERS} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{M} \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{M} \\ \text{CASE} & \text{NOM} \end{bmatrix} \end{bmatrix}$$

(234) S-structure for 'he'

$$\textit{he}: \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{M} \\ \text{HUMAN} & + \\ \text{ANIMATE} & + \end{bmatrix}$$

As can be seen in this example, the English pronoun 'he' has no mismatches either between INDEX and CONCORD or between the functional and semantic structures. The properties 'third person,

[134] For example, see Dalrymple, Lowe and Mycock (2019, 58–59): 'In many approaches, a word form which can have several values for a feature is treated by leaving the value for the feature underspecified or unspecified.'

singular and masculine' are represented at all three levels: INDEX, CONCORD and semantics. CASE is represented only as a feature within CONCORD (following standard definitions of INDEX and CONCORD feature bundles), and as it is a grammatical (rather than semantic) phenomenon, it is not included at the level of semantic structure. The semantic structure includes the purely semantic information that the referent is human and animate; the functional structure also provides the information that the PRED value is 'PRO', indicating that the form is a pronoun, and the feature PRONTYPE has the value PERS, indicating that it is a personal pronoun.[135]

It is worth briefly considering the link between the functional structure and semantic structure levels within the LFG architecture. The mapping between functional structure and semantic structure has been discussed by Asudeh and Giorgolo (2012), who discuss the role of the correspondence function σ[136] which maps functional structure to semantic structure, as originally proposed by Kaplan (1987) and developed by Dalrymple, Lamping and Saraswat (1993) and Butt, Dalrymple and Frank (1997). The focus of Asudeh and Giorgolo's paper is mainly on the inclusion of argument structure information within the semantic structure, which allows a simpler architecture and connected semantic structure than alternative proposals involving a separate argument structure (Asudeh and Giorgolo 2012, 8) (see brief discussion in section 4.4.10 above). They provide a clear example of the relationship between functional structure and semantic

[135] The role of the PRONTYPE feature is to give information about what type of pronoun the f-structure is describing, because there are various syntactic restrictions on which types of pronouns (e.g. personal pronouns PERS versus reflexive pronouns REFL versus relative pronouns REL) are needed in particular grammatical contexts (Dalrymple, Lowe, and Mycock 2019, 84).

[136] This is a different use of the symbol σ from its use as an abbreviation for 'a set of morphosyntactic features' in PFM (see section 2.2.1).

structure elements, given here as Figure 16 (Asudeh and Giorgolo 2012, fig. 2). The level of semantic structure represents participants as ARG1 and ARG2 rather than as thematic roles, using an event semantics (Parsons 1990) view of meaning in which 'thematic roles are functions from events to individuals' (Asudeh and Giorgolo 2012, 226–27).

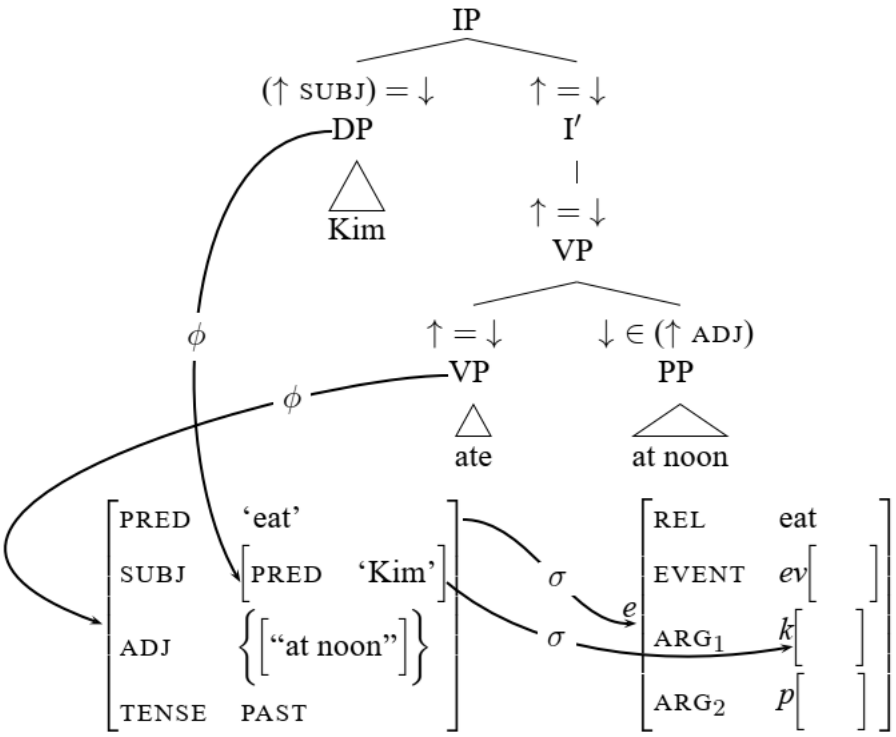
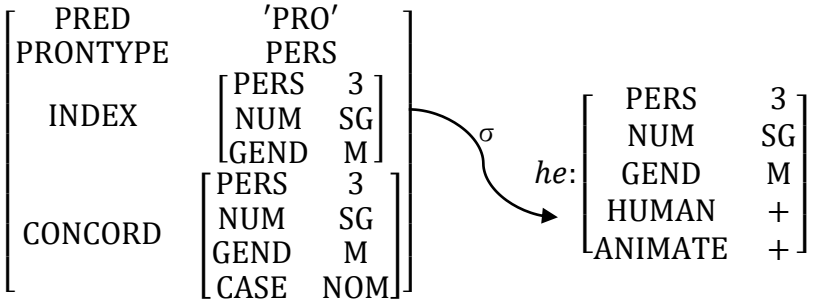


Figure 16 Analysis of 'Kim ate at noon'

The mapping function σ can also be applied to my functional and semantic structure representations of pronouns. I continue the example of the English pronoun 'he' above to indicate the σ mapping.

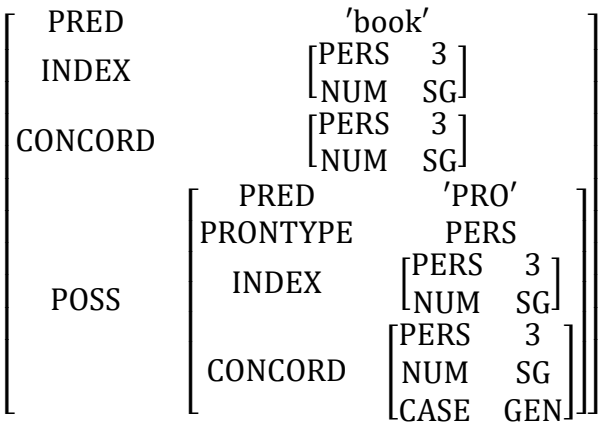
(235) F-structure and S-structure for 'he', with σ mapping



Since the mapping in this context is straightforward, I do not indicate it for the rest of the semantic and functional structures.

Next, I present another simple English example of the functional and semantic structures associated with the noun phrase 'his book'.

(236) F-structure for 'his book'



(237) S-structure for 'his book'

for 'book':

$$book: \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{HUMAN} & - \\ \text{ANIMATE} & - \end{bmatrix}$$

for possessor:

$$his: \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{M} \\ \text{HUMAN} & + \\ \text{ANIMATE} & + \end{bmatrix}$$

Here we see that the two structures capture information about both the possessor ('he') and the possessum ('book'). The functional structure captures information about the main predicate of the phrase, and about the grammatical features associated with the different elements of the noun phrase. We again see the PRED value PRO and the PRONTYPE PERS associated with the possessive pronoun 'his', which is represented in the structure as a POSS. As expected, the semantic structure represents the book's relevant semantic properties (singular, non-human and inanimate) and the possessor's, which are the same as in (234). I do not represent the feature 'case' for nouns in English (or Portuguese), as unlike pronouns, nouns do not have case distinctions.

Having seen the relevant features of functional and semantic structure, I now move on to give formal representations of each of the Portuguese grammaticalised pronouns.

6.4.1 Formal representations of second person singular pronouns

This section gives formal representations for the forms involved in second person singular pronouns, including the source noun phrase (*vossa mercê*), the inherited pronoun (*tu*) and the grammaticalised pronoun (*você*). The formal representations of *você* also reflect differences in the grammatical features of the pronoun as seen in different modern varieties of Portuguese. For each

form, I give a functional structure and a semantic structure, in line with the example given above in (233) and (234).

First, I consider the source noun phrase *vossa mercê*.

(238) F-structure for source noun phrase *vossa mercê* 'your mercy'

$$\begin{bmatrix} \text{PRED} & \text{'mercy'} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{F} \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{F} \end{bmatrix} \\ \text{POSS} & \begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{PERS} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & 2 \\ \text{NUM} & \textbf{PL} \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & 2 \\ \text{NUM} & \textbf{PL} \\ \text{CASE} & \text{GEN} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

(239) S-structure for *vossa mercê*

for 'mercy':

$$\textit{merc}\hat{\text{e}}\text{:} \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{HUMAN} & - \\ \text{ANIMATE} & - \end{bmatrix}$$

for possessor:

$$\textit{vossa}\text{:} \begin{bmatrix} \text{PERS} & 2 \\ \text{NUM} & \textbf{SG} \\ \text{GEND} & \\ \text{HUMAN} & + \\ \text{ANIMATE} & + \end{bmatrix}$$

The plurification strategy used for the possessor *vossa* 'your' is captured by the plural number feature for the possessor in the f-structure (238), which shows a mismatch with the singular number feature for the possessor in the s-structure (239).

I have not represented the gender of the noun *mercê* 'mercy' at the level of semantic structure because, as an abstract noun, this has no intrinsic gender. On this view, the syntactic gender (as represented at the levels INDEX and CONCORD) is unrelated to semantics. Instead it is a lexical property of the form, predictable from formal, morphophonological information (Corbett 1991). By contrast, the gender of the possessor is represented in the attribute-value matrix as underspecified, following the notational convention from f-structures where the underspecified feature is included but given a blank value (Dalrymple, Lowe, and Mycock 2019, 82).

We can make a similar analysis of the inherited second person singular pronoun *tu,* as found in standard Portuguese, using functional structure and semantic structure.

(240) F-structure for *tu* (inherited)

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{PERS} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & 2 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & 2 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \\ \text{CASE} & \text{NOM} \end{bmatrix} \end{bmatrix}$$

(241) S-structure for *tu* (inherited)

$$tu\text{:}\begin{bmatrix} \text{PERS} & 2 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \\ \text{HUMAN} & + \\ \text{ANIMATE} & + \end{bmatrix}$$

These representations capture the fact that there is no mismatch between the syntactic features (INDEX and CONCORD) of PERSON and NUMBER and the semantic features, for the inherited

pronoun. The pronoun has uniformly second person singular agreement on all targets.[137] Second person pronouns are not specified for gender in Portuguese, with the same form used for referents of either masculine or feminine gender, so the GENDER feature is left underspecified.

I now turn to the grammaticalised form *você*, used as a second person pronoun, which has different grammatical features, involving a mismatch between syntax and semantics. As we saw above, there are several different feature specifications for the functional structure of *você*, depending on its degree of grammaticalisation. In all cases, the pronoun will have the same semantic properties. The single S-structure presented in (242) is therefore applicable to *você* across all varieties of Portuguese: semantically second person and singular, human and animate.

(242) S-structure for *você* in European Portuguese

$$você: \begin{bmatrix} \text{PERS} & 2 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \\ \text{HUMAN} & + \\ \text{ANIMATE} & + \end{bmatrix}$$

The differences between the pronouns are made explicit at the level of functional structure. I present here (243)-(245) the three possible feature specifications required to account for the properties of the pronoun *você* in the varieties discussed above, then discuss each in turn.

[137] An exception to this is the Brazilian Portuguese varieties where *tu* is found with third person verb agreement; see Table 35 and footnote 116 above. As stated previously, this is likely due to reductions in person/number distinctions in the verbal inflectional paradigm leading to syncretism, and can be treated as a different phenomenon from the mismatches considered here.

(243) F-structure for *você* in European Portuguese

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{PERS} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & \mathbf{3} \\ \text{NUM} & \text{SG} \\ \text{GEND} & \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & \mathbf{3} \\ \text{NUM} & \text{SG} \\ \text{GEND} & \\ \text{CASE} & \text{NOM} \end{bmatrix} \end{bmatrix}$$

(244) F-structure for *você* in a less grammaticalised Brazilian Portuguese variety

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{PERS} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & \mathbf{3} \\ \text{NUM} & \text{SG} \\ \text{GEND} & \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & \mathbf{2} \\ \text{NUM} & \text{SG} \\ \text{GEND} & \\ \text{CASE} & \end{bmatrix} \end{bmatrix}$$

(245) F-structure for *você* in a more grammaticalised Brazilian Portuguese variety

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{PERS} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & \mathbf{2} \\ \text{NUM} & \text{SG} \\ \text{GEND} & \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & \mathbf{2} \\ \text{NUM} & \text{SG} \\ \text{GEND} & \\ \text{CASE} & \end{bmatrix} \end{bmatrix}$$

As discussed in section 6.2.1, in semantic terms, the syntactic behaviour of the European Portuguese form *você* (243) is uniformly third person. The form has third person agreement on all agreement targets (verbs, pronouns, possessives). For this reason, both the INDEX and CONCORD features must be third person. The number feature is singular throughout and the gender is not specified: it could have the value M or F depending on the addressee.

In varieties of Brazilian Portuguese close to the standard, *você* is grammatically identical to in European Portuguese. There are differences in the sociolinguistic implications associated with the form, including questions of politeness, formality, respect and so on, but given that there is as yet no general account of how to represent such sociolinguistic or pragmatic factors in LFG, I do not include such features in the representations given here.

In the intermediate (less grammaticalised) variety of Brazilian Portuguese, illustrated in (244), we see a mismatch at the level of functional structure. Verb agreement, which is controlled by INDEX (Wechsler and Zlatić 2003, 14), is third person, but other targets show second person agreement (see for example (201) and (202) in section 6.2.1.3). This must be represented via the value 2 for the CONCORD PERSON feature. My inclusion of PERSON information in the CONCORD set of features is an innovation introduced in order to capture this kind of mismatch.[138]

---

[138] An alternative analysis may also be available which removes the need to include PERSON in CONCORD, but which is otherwise compatible with my analysis. On this view, all instances of syntactic agreement are instances of INDEX agreement, while instances of semantic agreement (analysed here as evidence for the PERSON feature existing in CONCORD) in fact refer directly to the semantic features, rather than to the syntax. In one analysis by Wechsler (2011, 1008–21), semantic agreement results when the controller is not specified for a particular INDEX feature. This implies that grammaticalised pronouns in different Portuguese varieties vary as to whether they have particular features (as expressed on different targets), which can be straightforwardly implemented at the level of the lexical entry in an LFG setting. More broadly, applying this analysis within LFG would involve altering the features proposed for the pronouns at the levels of semantic structure and functional structure, with the σ projection also being involved. This proposal merits investigation in relation to Portuguese, but is not developed further in this thesis.

In addition, the pronoun *você* in this variety can be used as an oblique pronoun as well as a (nominative) subject, and is therefore underspecified for the feature CASE.

In the more grammaticalised Brazilian Portuguese variety, shown in (245), the form *você* has uniformly second person agreement on all targets. Unlike in other Portuguese varieties described above, the PERSON feature on all targets is second person (see example (202) and surrounding discussion in section 6.2.1.3). The INDEX and CONCORD PERSON features are therefore both 2, here matching the semantic information. This reflects my analysis that in a fully grammaticalised form, there is no mismatch between the semantic features of the referent and the syntactic features of the corresponding pronoun. The pronoun *você* in this variety is also underspecified for CASE.

6.4.2 Formal representations of second person plural pronouns

In this section, I present equivalent analyses and discussion of the second person plural pronouns. I first consider the inherited second person plural form *vós*, before moving on to present the grammaticalised form *vocês*. Again, we see that the mismatches between the semantics and syntax of *vocês* at various stages of its grammaticalisation are clearly represented in the formal analyses.

First, I present the inherited second person plural form *vós*:

(246) F-structure for *vós* (inherited)

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{PERS} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & 2 \\ \text{NUM} & \text{PL} \\ \text{GEND} & \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & 2 \\ \text{NUM} & \text{PL} \\ \text{GEND} & \\ \text{CASE} & \text{NOM} \end{bmatrix} \end{bmatrix}$$

(247) S-structure for *vós* (inherited)

$$vocês: \begin{bmatrix} \text{PERS} & 2 \\ \text{NUM} & \text{PL} \\ \text{GEND} & \\ \text{HUMAN} & + \\ \text{ANIMATE} & + \end{bmatrix}$$

As we saw for the inherited form *tu* above, there is no mismatch within the syntax, or between the syntax and semantics.

Now I turn to the grammaticalised second person plural pronoun *vocês*. Regardless of its level of grammaticalisation, this form will have a semantic structure like that of the inherited second person plural form *vós*:

(248) S-structure for *vocês*

$$vocês: \begin{bmatrix} \text{PERS} & \mathbf{2} \\ \text{NUM} & \text{PL} \\ \text{GEND} & \\ \text{HUMAN} & + \\ \text{ANIMATE} & + \end{bmatrix}$$

Again, we see several different functional structures associated with different degrees of grammaticalisation.

(249) F-structure for *vocês* (standard variety)

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{PERS} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & \mathbf{3} \\ \text{NUM} & \text{PL} \\ \text{GEND} & \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & \mathbf{3} \\ \text{NUM} & \text{PL} \\ \text{GEND} & \\ \text{CASE} & \text{NOM} \end{bmatrix} \end{bmatrix}$$

(250) F-structure for *vocês* in varieties with more semantic agreement

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{PERS} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & \mathbf{3} \\ \text{NUM} & \text{PL} \\ \text{GEND} & \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & \mathbf{2} \\ \text{NUM} & \text{PL} \\ \text{GEND} & \\ \text{CASE} & \text{NOM} \end{bmatrix} \end{bmatrix}$$

(251) F-structure for Brazilian Portuguese *vocês* with underspecified case

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{PERS} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{PL} \\ \text{GEND} & \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{PL} \\ \text{GEND} & \\ \text{CASE} & \end{bmatrix} \end{bmatrix}$$

In the standard variety of Portuguese (249), *vocês* has uniformly third person plural agreement, on all agreement targets. For this reason, both the INDEX and CONCORD features for PERSON and NUMBER are specified as 3 and PL respectively. There is therefore a mismatch between the PERSON feature at functional structure (third person) and the corresponding feature at semantic structure (second person).

As discussed in section 6.2.1.4, in some European Portuguese varieties we find *vocês* with the inherited second person plural possessives and clitics (e.g. *vosso*, *vos*), indicating that it has the value 2 for PERSON in CONCORD, as shown in (250). This allows agreement targets controlled by CONCORD features (such as the possessive and the object pronoun) to be second person plural in form (*vosso* 'your.PL'). On the other hand, the INDEX PERSON feature remains 3. We never get a 'fully grammaticalised' form of *vocês* where INDEX agreement is second person plural, which would give us second person plural verb forms as well as possessives, etc. As noted above, there is no

evidence of the grammaticalised pronoun *vocês* being used with second person plural verb agreement (controlled by INDEX).

We might ask why there is no evidence of *vocês* taking second person plural verbs, given that by contrast we do find second person singular verbs with *você* in some Brazilian Portuguese varieties. One possible explanation is that the inherited form *tu* has been more resilient across Portuguese varieties than *vós*, so many speakers who are not familiar with the verb paradigm for second person plural forms are used to using the second person singular forms, allowing occasional use of *você* with second person singular forms. By contrast, we have seen that other varieties of Brazilian Portuguese use the inherited second person singular pronoun *tu* with third person singular verb forms, implying that similar morphosyntactic changes can be led either by the loss of forms from the verb paradigm, or by loss of a pronoun with its associated forms. Further discussions of the interactions between the rise of grammaticalised pronoun forms and the reduction of person/number distinctions in the verbal paradigm can be found in Zilles (2005), Aalberse and Stroop (2015) and Lara Bermejo and Guilherme (2018). While Aalberse and Stroop argue that the loss of inflectional person/number distinctions led to the rise of grammaticalised pronouns, Lara Bermejo and Guilherme argue that in fact the process is led by changes in the pronoun system, with the loss of verbal inflection being a consequence of the decreased use of the person/number distinctions because the grammaticalised pronouns do not require them (2018, 352).

Another variant of *vocês*, found in Brazilian Portuguese, takes third person plural verb agreement, but is underspecified for case (acting as subject, direct object and oblique). In this instance, its CONCORD PERSON feature is unclear, because person is not marked on coreferent pronouns, with the possessive being replaced by '*de vocês*', the direct object by '*vocês*' and the indirect object by '*para vocês*'. This is shown in (251).

Having seen the analysis of second person plural pronouns, I now move onto the first person plural pronoun.

6.4.3 Formal representations of first person plural pronouns

This section gives representations of a formal analysis of the grammaticalised first person plural form *a gente*, including functional structures and semantic structures. These include complexities similar to those seen in the second person. In the representations of *a gente* at different stages of grammaticalisation, we again see that the syntax of the form moves away from the syntactic features of the original noun phrase (singular, third person and feminine,) towards the semantic content associated with the meaning (plural, first person and underspecified for gender). This move from inherited syntactic features to semantically aligned features accords with a more general view of the grammaticalisation of pronouns from nouns, which 'tend to lose morphological distinctions of number, gender, case, definiteness … etc.' (Heine and Song 2011, 593).

As we saw in previous examples, the PRED value of the form moves from being contentful (having the value 'people') to being anaphoric (having the value PRO).[139]

First, I give the f-structure and s-structure for the original noun phrase *a gente*.

(252) F-structure for source noun phrase *a gente* 'the people'

$$\begin{bmatrix} \text{PRED} & \text{'people'} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{F} \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{F} \end{bmatrix} \\ \text{DEF} & + \end{bmatrix}$$

[139] The form *a gente* has not been phonetically eroded to the extent of *você*. It is beyond the scope of my analysis to present details about phonetic change, so this information is not represented at the levels included here.

(253) S-structure for *gente* 'people':

$$gente: \begin{bmatrix} \text{NUM} & \text{SG} \\ \text{GEND} & \\ \text{HUMAN} & + \\ \text{ANIMATE} & + \end{bmatrix}$$

I have not included a semantic structure for the definite article *a* 'the'. The correct semantic analysis of the definite article is somewhat complex and a solution is beyond the scope of this thesis. However, the semantic structure will include an indication of definiteness.

The gender of the noun *gente* 'people' is underspecified at the level of semantic structure because the noun can refer to people of either gender or a mixed group.[140] However, as we have seen, syntactic gender (as represented at the levels INDEX and CONCORD) is independent of semantic gender, instead being a lexical property of the form: the syntactic gender of the noun (feminine) is represented at the level of functional structure.

As a point of comparison, I also give here a functional structure and semantic structure for the inherited first person plural pronoun *nós*, which are straightforward, involving no mismatches.

(254) F-structure for inherited form *nós*

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{PERS} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & 1 \\ \text{NUM} & \text{PL} \\ \text{GEND} & \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & 1 \\ \text{NUM} & \text{PL} \\ \text{GEND} & \\ \text{CASE} & \text{NOM} \end{bmatrix} \end{bmatrix}$$

[140] This is also the correct analysis of the semantic gender of *a gente* as a pronoun.

(255) S-structure for inherited form *nós*

$$nós: \begin{bmatrix} \text{PERS} & 1 \\ \text{NUM} & \text{PL} \\ \text{GEND} & \\ \text{HUMAN} & + \\ \text{ANIMATE} & + \end{bmatrix}$$

We saw in section 6.2.2 that *a gente* moves from being a noun phrase to being a generic pronoun, before it becomes a personal pronoun. This is represented in (256) and (257).

(256) F-structure for *a gente* as generic pronoun

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{GENERIC} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{F} \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{F} \\ \text{CASE} & \end{bmatrix} \end{bmatrix}$$

(257) S-structure for *a gente* as generic pronoun

$$a\ gente: \begin{bmatrix} \text{PERS} & \\ \text{NUM} & \\ \text{GEND} & \\ \text{HUMAN} & + \\ \text{ANIMATE} & + \end{bmatrix}$$

In these representations, we see that the PRONTYPE feature has the value GENERIC, representing the form's status as a generic pronoun. However, the other syntactic features of the form at the level of functional structure remain as they were in the source form: feminine, singular and third

person.[141] Semantically, the form's number is unclear, since it has properties similar to a mass noun, not being clearly countable, and therefore is represented here as underspecified. The appropriate value of the semantic PERSON feature is also represented as underspecified, because in its generic use, the pronoun refers to any group of people, not necessarily excluding conversational participants (first and second person). The GENDER feature is also underspecified. The form is still specified as being +HUMAN and +ANIMATE.

In the next stage of grammaticalisation, *a gente* loses some of the grammatical features associated with the source noun phrase, namely the feminine gender, and according to Lopes (2003, 80–81), the third person. At this point, it still functions as a generic pronoun, meaning that in practice, it occurs with third person verbs.

(258) F-structure for generic *a gente*, some features lost:

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{GENERIC} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & \\ \text{NUM} & \text{SG} \\ \text{GEND} & \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & \\ \text{NUM} & \text{SG} \\ \text{GEND} & \\ \text{CASE} & \end{bmatrix} \end{bmatrix}$$

Having lost its third person feature, the pronoun gains an alternative function as a personal pronoun for the first person plural, from the 1900s onwards (C. R. dos S. Lopes 2003, 80–83). Its

[141] The case to be given to such a form is somewhat unclear. It can act as a subject or oblique form, with similar behaviour to a normal Portuguese noun phrase. For this reason I have represented the case as underspecified here.

semantic structure at this stage (259) includes the feature PERSON: 1. Again, this semantic structure is valid for all variants of *a gente* as a first person plural pronoun, whatever its grammatical features.

(259) S-structure for *a gente* with syntactic agreement

$$a\ gente: \begin{bmatrix} \text{PERS} & 1 \\ \text{NUM} & \text{PL} \\ \text{GEND} & \\ \text{HUMAN} & + \\ \text{ANIMATE} & + \end{bmatrix}$$

The grammaticalised personal pronoun has the same grammatical features as the generic pronoun, although its pronoun type is now represented as PERS:

(260) F-structure for *a gente* with syntactic agreement

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{PERS} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & \\ \text{NUM} & \text{SG} \\ \text{GEND} & \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & \\ \text{NUM} & \text{SG} \\ \text{GEND} & \\ \text{CASE} & \end{bmatrix} \end{bmatrix}$$

In the next stage, we see some further changes in the features of the pronoun, with the PERSON and NUMBER features in CONCORD being specified as first person and plural, as the application of the pronoun becomes more restricted. Verb agreement (which is controlled by INDEX) remains third person and singular, but other agreement targets, such as possessives, (which are controlled by CONCORD) have first person plural features. This pattern of mixed agreement is quite common in Brazilian Portuguese varieties, according to Lopes (2003) (see section 6.2.2), with examples such as (213), repeated here for convenience as (261).

(261)

a gente	cria	nosso	próprio	universo
A GENTE	create.3SG	1PL.POSS	own.MSG	universe.MSG
com	o	poder	da	mente
with	the.MSG	power.MSG	of.the.FSG	mind.FSG

'We create our own universe with the power of the mind.'

(*Corpus do Português*, Web/Dialects)

The features of this stage in grammaticalisation are summarised in (262).

(262) F-structure for *a gente* with mixed agreement

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{PERS} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & 3 \\ \text{NUM} & \text{SG} \\ \text{GEND} & \text{F} \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & 1 \\ \text{NUM} & \text{PL} \\ \text{GEND} & \\ \text{CASE} & \end{bmatrix} \end{bmatrix}$$

The final possible stage of grammaticalisation is for verb agreement to also be first person plural. This would be represented formally through both INDEX and CONCORD features being first person and plural, thereby aligning with the semantics of the pronoun. Again, we see a gradual move from syntactic to semantic agreement, in line with the agreement hierarchy (section 6.1.2).

There is some marginal evidence of this pattern occurring in non-standard Brazilian Portuguese, as we saw in example (218), repeated here as (263):

(263)

essas	11	hrs (= horas)	a gente	vamos	ter	que
those	11	hours	A GENTE	go.PRES.1PL	have.INF	that
trabalhar	como	loucos				
work.INF	like	mad.MPL				

'We're going to have to work like mad for those 11 hours'

(*Corpus do Português*, Web/Dialects)

However, this pattern is unusual, even in non-standard varieties of Portuguese, possibly because of general pressures to reduce the person/number distinctions found in the verbal paradigm (see discussion in section 6.4.2).

If a full transition to semantic agreement (including verb agreement) is realised, the features of the grammaticalised form *a gente*, as seen in (263), can be represented as below.

Analysis for semantic agreement (most fully grammaticalised)

(264) F-structure for *a gente* with semantic agreement

$$\begin{bmatrix} \text{PRED} & \text{'PRO'} \\ \text{PRONTYPE} & \text{PERS} \\ \text{INDEX} & \begin{bmatrix} \text{PERS} & 1 \\ \text{NUM} & \text{PL} \\ \text{GEND} & \end{bmatrix} \\ \text{CONCORD} & \begin{bmatrix} \text{PERS} & 1 \\ \text{NUM} & \text{PL} \\ \text{GEND} & \\ \text{CASE} & \text{NOM} \end{bmatrix} \end{bmatrix}$$

Unlike in the less grammaticalised instances, here the grammaticalised pronoun has uniformly first person and plural agreement features. The INDEX and CONCORD PERSON features are therefore both 1, and the NUMBER features are PL, here matching the semantic information, as shown in (259).

For *vocês* above, we saw that there was no evidence of the grammaticalised pronoun being used with second person plural verb forms, perhaps because those verb forms had already been eliminated from the paradigm when the pronoun was at an earlier stage of grammaticalisation. The interaction of the loss of particular person/number combinations for verb forms (in this case, first person plural) and the retention or loss of equivalent forms for other agreement targets such as possessives, object pronouns, is an interesting area of research, as was summarised in section 6.4.2.

6.4.4 Lexical and lexemic entries for inherited and grammaticalised pronouns

Both lexical and lexemic entries play an important part in the interaction of the lexicon, the morphological component and the syntax, as we saw in Chapter 2.

To remind the reader of these concepts, the lexical entry $\mathcal{L}$ is a relation between an inflected wordform and other pieces of information, taking the form of a four-place relation:

(265) $\mathcal{L}$<s-form, p-form, category, f-description>

The s-form is 'the lexical representation specifying the syntactic string associated with a word' and the p-form is 'the lexical representation specifying the prosodic string contributed by that word' (Dalrymple, Lowe, and Mycock 2019, 437). The f-description contains syntactic, semantic and other types of information, including, for example, meaning constructors associated with the word and information-structural properties, which can be captured in templates (Dalrymple 2015, 3–4).

On the other hand, the lexemic entry is a three-place relation that contains information relevant to the lexeme as a whole: its idiosyncratic morphological forms and features, its functional description, and its Lexemic Index (LI). Again, the functional description will include syntactic, semantic and other information associated with the lexeme as a whole.

(266) LE < root/idiosyncratic stem forms/features, f-description, LI >

Sample lexical entries for a lexical item *bolo* 'cake' and a functional item *o* 'the.MSG' are given in (267) and (268), including meaning constructor labels as a representation of semantic information (as discussed in 3.2.1).

(267) bolo: $\mathcal{L}$<bolo, /bolu/, N,

{(↑PRED)='CAKE', (↑INDEX GEND)=MASC, (↑INDEX NUM)=SG

[**bolo**]} >

(268) o: $\mathcal{L}$< o, /u/, Det,

{(↑DEF)=+, (↑INDEX GEND)=MASC, (↑INDEX NUM)=SG

[**o**]}>

In general, CONCORD features are inherited by default from the INDEX features, so do not need to be specified in the lexical entry (as shown in Dalrymple, Lowe, and Mycock 2019, 438–39). Where there is a mismatch between CONCORD and INDEX features, both features can be included in the lexical entry.

The lexemic entry includes only the f-description information associated with the lexeme as a whole, which often has substantial overlap with the f-description for the wordform. For example, the lexemic entries for the items given above would include the following information:[142]

[142] Since I have not included a full analysis of inflection classes for nominal forms or the definite article, I give these an arbitrary inflection class (M-ICLASS) value of 1. I also give the forms a lexemic index consisting of their orthographic form and the letter 1.

(269) Bolo: LE < {M-STEMROOT: bolo, M-ICLASS:1, M-CAT:NOUN},
{(↑PRED) = 'CAKE', (↑INDEX GEND) = MASC
[**bolo**]}
BOLO1>

(270) o: LE < {M-STEMROOT: o, M-ICLASS:1, M-CAT:DET},
{(↑DEF)=+, (↑INDEX GEND)=MASC
[**o**]}>
O1>

The precise distribution of f-description information between the lexical and lexemic entries is debatable, particularly for the definite article *o* 'the'. It would be possible to treat all person/number specifications of the definite article as the same lexical item, in which case the lexemic entry would not include a specification of masculine gender. In my example lexemic entry for this word, I have included the masculine gender, but not a number feature, implying that the definite article can be inflected for singular or plural number (singular number is given in the lexical entry for the word, in (268)).

The lexical and lexemic entries for Portuguese pronouns take similar forms to those of other Portuguese words, as we would expect. In each case, the relevant PERSON, NUMBER, GENDER and CASE features, for INDEX and CONCORD, must be specified in the f-description. When the CONCORD features are predictable from the INDEX features, they can be omitted. When a form is underspecified for a particular feature, this is reflected by the feature having no value after the equals sign, similar to the convention for representing underspecified features at f-structure by including the feature with a blank value (Dalrymple, Lowe, and Mycock 2019, 82).

The most relevant parts of the lexical and lexemic entries are the f-descriptions, which reflect information about grammatical features that is represented in the functional structure. For this

reason, it is not necessary to work through the lexical and lexemic entries for all the possible variants of all the pronouns in full here, but I present some key examples.

First, we can consider the inherited second person singular form *tu*, which has no mismatches.

(271) tu: ℒ< tu, /tu/, Pro,

{(↑DEF)=+, (↑INDEX PERS)=2, (↑INDEX NUM)=SG

[**tu**]} >

As an example of a pronoun with mismatches at the level of functional structure, we could consider *você* as found in the 'less grammaticalised' Brazilian Portuguese variety. For this form, the PERSON feature in INDEX is mismatched with that in CONCORD, so both must be specified in the lexical entry. The NUMBER feature, however, is singular in both INDEX and CONCORD, so the CONCORD NUMBER feature is not specified and can be inherited from INDEX.

(272) você: ℒ< você, / vɔ.ˈse/, Pro,

{(↑DEF)=+, (↑INDEX PERS)=3, (↑CONCORD PERS)=2, (↑INDEX NUM)=SG

[**você**]} >

The lexemic entries for these two forms (inherited *tu* and grammaticalised *você* with a mismatch at functional structure) are similar, with some differences in the morphological form and properties, and grammatical features.

(273) tu: LE < {M-STEMROOT: tu, M-ICLASS:1, M-CAT:NOUN},

{(↑PRED) = 'PRO', (↑INDEX PERS) = 2, (↑INDEX NUM) = SG, (↑INDEX GEND)

=

[**tu**]}

TU1>

(274) você: LE < {M-STEMROOT: você, M-ICLASS:1, M-CAT:NOUN},
{(↑PRED) = 'PRO', (↑INDEX PERS) = 3, (↑CONCORD PERS)=2 (↑INDEX NUM)
= SG, (↑INDEX GEND) =
[**você**]}
VOCÊ1>

We can also see an example with a mismatch between the functional information and the semantic information, such as standard *você* which has uniformly third person grammatical features but second person semantics. In the lexemic entries given here, information about semantics is abbreviated by the label for the meaning constructor [**você**], which stands in place of a fuller analysis of the semantics conveyed by the pronoun, including the fact it has second person reference. The mismatch between third person functional information and second person semantics is therefore captured at the level of the lexemic entry.

(275) você: LE < {M-STEMROOT: você, M-ICLASS:1, M-CAT:NOUN},
{(↑PRED) = 'PRO', (↑INDEX PERS) = 3, (↑INDEX NUM) = SG, (↑INDEX GEND)
=
[**você**]}
VOCÊ1>

The lexemic and lexical entries associated with each pronoun therefore bring together all the information of various different types that is necessary to account for the form's behaviour, in terms of its syntax, morphology, phonology and meaning.[143] Mismatches in features at different

[143] To provide an analysis of the anaphoric reference properties of pronouns, we also require an account of discourse structure. LFG has a well-established and fully explicit account of anaphoric reference, building on the glue semantics approach, as set out for example by Dalrymple, Lowe

levels are represented in the lexicon, which is necessary because these mismatches are idiosyncratic and unpredictable, affecting only individual lexical items.

The final element of my representation of the mismatch behaviour of the Portuguese grammaticalised pronouns is the PFM rules that account for the morphological forms. These are presented in the next section.

6.4.5 PFM rules for pronouns

The PFM rules required to account for Portuguese pronouns are relatively simple. As we saw above, pronouns have very small inflectional paradigms, including only different case forms. However, many pronouns do not have distinct case forms. Each person/number combination is represented as a different lexical item, and the inherited pronouns are (of course) separate lexical items from the grammaticalised pronouns.

An example of a pronoun that inflects for different cases is the first person pronoun *eu* 'I', which has the oblique form *mim* and the possessive form *meu*.[144] It also has a clitic form *me* which acts as direct object, indirect object and reflexive. Although clitic forms can be generated as verbal affixes (see discussion of Luís's (2003) proposals above in section 6.3.2), for the sake of completeness I include them here.

---

and Mycock (2019, sec. 14.4). This typically involves a version of Discourse Representation Theory (DRT) (Kamp [1981] 2002), such as Partial Compositional Discourse Representation Theory (Haug 2014; Dalrymple, Haug, and Lowe 2018).

[144] This is the possessive form used with a masculine singular possessum. The forms for feminine singular, masculine plural and feminine plural are respectively *minha, meus, minhas*.

All of these forms must be specified in the morphological rules, since they are not the product of regular inflectional rules for nouns. (I have arbitrarily started this block of realisation rules for nominal forms at 301, so that the numbering does not overlap with any of the rules for verbs presented in previous chapters.)

301. RR_1,{case: nom}, (eu) ∩ **N** ∩ **Pt**(<χ, σ>) $=_{def}$ <[χ_i], σ> for all i

302. RR_1,{case: obl}, (eu) ∩ **N** ∩ **Pt**(<χ, σ>) $=_{def}$ <[mim], σ> for all i

303. RR_1,{case: poss}(eu) ∩ **N** ∩ **Pt**(<χ, σ>) $=_{def}$ <[meu], σ> for all i

304. RR_1,{case: dobj|indobj|refl}(eu) ∩ **N** ∩ **Pt**(<χ, σ>) $=_{def}$ <[me], σ> for all i

Similar rules will apply for the formation of other inherited pronouns which have a complete inflectional paradigm of case forms.

The grammaticalised pronoun forms have an even more straightforward morphology, being morphologically inert. For example, the rule to account for the form *você* which can act as a nominative or an oblique is similar to the rule for the subject (nominative) form of *eu*:

305. RR_1,{case: nom|obl}, (você) ∩ **N** ∩ **Pt**(<χ, σ>) = $_{def}$ <[χ_i], σ> for all i

Similar rules will also apply to capture the morphology of the grammaticalised forms *vocês* and *a gente*.

In general, the PFM rules required for pronouns present few complexities. Each rule is extremely narrow, only applying to one input form, to ensure that it applies only to the relevant lexical item. Few inflectional processes are captured in the rules of this rule block.

Having seen the different parts of a LFG-PFM analysis of the Portuguese grammaticalised pronouns, I now move on to consider the value of this analysis in giving new insights, as well as summarising the adaptations to the framework I have proposed here.

6.5 Analysis: LFG-PFM insights into grammaticalised pronouns

As we have seen in the previous sections, LFG functional structures and semantic structures can be used to represent fine-grained detail about forms undergoing grammaticalisation. In particular, these structures are useful for representing two types of mismatches: mismatches between the forms' semantic value and syntactic behaviour, and mismatches within functional structure. The grammaticalised pronoun forms presented in this chapter are of particular interest as mismatch case studies due to the complexity of their grammatical features.

In this section, I first discuss adaptations I have made to the theories to provide a full analysis of the Portuguese forms. Next, I consider the value of using LFG and PFM to capture the morphological and syntactic properties of forms that have agreement mismatches, both looking at the Portuguese case studies, and more generally. I also mention questions raised by these analyses, and possible areas for development in future work.

6.5.1 Adaptations to LFG representations

I have proposed two adaptations to the current standard representation of features in LFG. The first relates to the properties included at the level of functional structure. My analyses include PERSON in the CONCORD bundle of grammatical features, in order to capture mismatch behaviour in the person feature, as we saw for *você* in some Brazilian Portuguese varieties, which have second person agreement on some targets and third person on others. In previous analyses, only NUMBER and GENDER appear in both INDEX and CONCORD, likely because this work focuses on mismatches in number and gender, such as the Serbian-Croatian noun *deca* 'children' discussed in section 6.1.2. Considering the development of personal pronouns and their grammatical feature behaviour has made clear the need for the expansion of the set of features in CONCORD to include PERSON. This proposal contradicts the standard claim that PERSON features cannot be part of CONCORD, because the diachronic source of CONCORD features is that they are incorporated noun classifiers (which would not include PERSON) while INDEX features are from incorporated pronouns (Wechsler 2011,

1017). My work contributes to this debate: any effective analysis of the person mismatches found in *você* must incorporate a mechanism similar to that proposed here, with the possibility to specify two different person features in the syntactic information for such forms. Within LFG the obvious solution is to expand the existing mechanism of INDEX and CONCORD that already accounts for similar mismatches involving number and gender. Further evidence in support of these proposals might be found by investigating the properties of person mismatches in other languages, such as imposters (Collins and Postal 2012) to establish whether these can be accounted for by INDEX and CONCORD in the way proposed here for the Portuguese person mismatches.

A second innovation made here to standard LFG representations is the inclusion of PERSON, NUMBER and GENDER features at the level of semantic structure. The addition of information to the s-structure is not unprecedented, having been previously proposed for the features ±ANIMATE and ±HUMAN, among others (Liao 2010). However, the inclusion of this information in particular is novel. Similar to the inclusion of the PERSON feature in CONCORD, this could be perceived as an unnecessary proliferation of similar information in different parts of the grammatical system. However, in this case too, there is a clear argument for including information about the real-world referent's properties in the semantics. Including these features in the semantic structure allows us to capture aspects of grammatical change that would otherwise go unexplained, namely the gradual development from so-called 'syntactic agreement' to 'semantic agreement'. As we have seen in the grammaticalised pronouns, the syntactic features of PERSON, NUMBER and GENDER in both INDEX and CONCORD, which start off as grammatical features matching the source noun phrase, may eventually come to match the corresponding semantic properties of the referent of the pronoun, although such a path of development may stop at any point.

6.5.2 Value of an LFG-PFM approach to mismatches

Throughout the thesis, I have shown that combining LFG with PFM provides a powerful tool for analysing the fine-grained changes that make up diachronic developments. This is particularly clear

in the example of the grammaticalised pronouns in Portuguese, where we see many instances of mismatches between different levels of grammatical structure, particularly functional and semantic structure. These mismatches vary across different Portuguese varieties, both historically and between distinct modern varieties. One of the key strengths of the LFG-PFM approach is the ability to separate out these different levels and capture the mismatches between them, explaining why certain mismatches arise and clarifying possible future directions for grammatical change that can eliminate (or maintain) mismatches between semantics and syntax, or within syntax.

As we have seen in this chapter, the INDEX and CONCORD feature bundles are an effective tool in the analysis of complex agreement behaviour, including mismatches. The values of corresponding features in the INDEX and CONCORD can change independently of each other, and can also change independently of the corresponding feature at the level of semantic structure, and independently of the corresponding morphological features. The necessary separation between syntactic features and morphological features has long been understood in LFG. Implementing PFM as the morphological component (which refers only to morphological features, not to syntactic features) clarifies the different roles of these sets of features within the architecture. We thus separate out four sets of closely related but essentially independent features: semantic features, INDEX features, CONCORD features and morphological features. This allows us to capture the hybrid agreement behaviour of the grammaticalised pronouns in synchronic terms, as well as allowing a precise account of the diachronic development of the forms from their source noun phrases.

Some information about mismatches within the pronouns' grammatical and semantic properties is encoded in the pronouns' lexemic entries, which capture the relationship between particular forms, their meanings, and their grammatical properties. As we have seen, the lexemic entry and lexical entry provide the information needed for the morphological component. Even though pronouns have small paradigms in Portuguese, the PFM machinery allows a simple and effective account of the morphological properties of the different forms.

A crucial part of the grammaticalisation process from the source noun phrases to their eventual status as pronouns is a change in the PRED feature, which captures the reference of the phrase. While the source noun phrases have a contentful and independent semantic PRED value, the grammaticalised pronouns have the value PRO, as discussed in section 6.3.1.2 above. This reflects the reduction in lexical information carried by the pronouns, which only have anaphoric reference, compared with full noun phrases, which have independent reference, captured by a lexical PRED value. This represents the start of the process of desemanticisation (semantic bleaching), a characteristic of grammaticalisation.

The LFG functional structure also allows us to capture other aspects of diachronic change in the grammaticalisation of pronouns from nouns. We saw that the form *a gente* first grammaticalised from the source noun phrase as a generic pronoun, which later extended in use to become a first person plural pronoun. This is reflected in the functional structure by a change in the value of the feature PRONTYPE from INDEF to PERS. Existing LFG work has not considered diachronic change in PRONTYPE values, as far as I am aware: however, the Portuguese data suggest there may be interesting patterns to describe within or across languages in terms of which changes are possible or likely.

Finally, I have demonstrated the usefulness of the LFG-PFM approach in accounting for historical processes. The structures presented here represent not only the start and end points of the grammaticalisation process (so far), but also exemplify stages in development. LFG representations of the kind given here are appropriate for fine-grained representations of diachrony: the incremental nature of grammatical change can be represented by single adjustments at the level of f-structure and/or s-structure.

6.5.3 Questions raised and areas for future research

The analysis given in this chapter has demonstrated that the LFG-PFM model allows an effective, explicit account of the different changes that can take place as noun phrases grammaticalise into

pronouns. Seeing this demonstration of the model for the Portuguese pronouns discussed here, *você*, *vocês* and *a gente*, also raises some further questions and areas for future research, in relation to the theoretical architecture and the Portuguese case studies.

One question raised by this case study of grammaticalised pronouns is whether there are other, similar mismatches in Portuguese noun phrases. We have seen in the pronouns that we find two specifications of mismatches, one where the pronoun's grammatical properties are split from its semantics (276) and one where the concord and index features are split, with semantics aligning with index (277). These occur in addition to the many forms in Portuguese that have no mismatches (278), including the 'fully grammaticalised' pronouns that are found marginally in Brazilian Portuguese.

(276) DECLENSION ↔ CONCORD ↔ INDEX | SEMANTICS

(277) DECLENSION ↔ CONCORD | INDEX ↔ SEMANTICS

(278) DECLENSION ↔ CONCORD ↔ INDEX ↔ SEMANTICS

An obvious question is what other types of mismatches are found in Portuguese nouns. Wechsler and Zlatić (2000, 15–17) discuss mismatches between DECLENSION and CONCORD in Serbo-Croatian, giving examples such as *kolega* 'colleague' which declines like a feminine noun, but has masculine INDEX and CONCORD features (expressed on all agreement targets). The question of classifying Portuguese nouns into declension classes is somewhat complex, but there are clear morphological patterns in gender and number inflection, as any description of Portuguese indicates (e.g. Dubert and Galves 2016, 419–20). On a simple view, nouns ending in *-a* tend to be feminine, so the masculine Portuguese noun (*o*) *colega* '(the.MSG) colleague' may be an example of this kind of mismatch. Further investigation will clarify how common the different types of mismatch are in Portuguese, providing further justification for Wechsler and Zlatić's four-way distinction between features, as used in and captured by my analysis.

The mismatches seen in the pronouns discussed in this chapter occur between INDEX and CONCORD features at the level of functional structure, or between grammatical features and semantic features, that is, at the levels of functional and semantic structure. By contrast, the double past participle mismatches discussed in Chapter 4 and Chapter 5 occurred at the interface of morphology with syntax, particularly in the relationship of the voice feature to the selection of a particular morphological form. The diverging historical development of the different phenomena is represented at different levels of representation, with the verbal developments mainly represented in the PFM machinery (the morphological component), and the noun phrase/pronominal developments represented in the functional and semantic structures of the LFG architecture. One question this raises is what other kinds of mismatches and divergences take place in Portuguese (or beyond) at these different levels of structure. For example, do we also find mismatches in functional and semantic structure for verb phrases, or mismatches in the morphology of noun phrases? Answering these questions could lead to a typology of possible changes seen in Portuguese and cross-linguistically, and shed more light on possible patterns of historical change. For example, we might find that particular parts of the grammar are more likely to show particular types of mismatches, or we might find trends across certain language families.

The architecture and types of analysis demonstrated here for the Portuguese grammaticalised pronouns could also be extended to other Romance varieties with grammaticalised pronouns, and to similar person/number/gender mismatches in and beyond Romance varieties. One example is studying the development of the polite personal pronoun *maga* from a reflexive pronoun meaning 'self' in Hungarian (de Groot 2009, 1010–11), which can be represented as a change in the feature PRONTYPE from the value REFL to the value PERS, accounting for changes in the grammatical contexts in which the pronoun can be used, along with changes in semantic reference captured at the level of semantic structure. Of course, such developments are far from limited to the grammaticalisation of personal pronouns. Similar types of changes may also fruitfully be investigated using the LFG architecture links between functional and semantic features.

Another important question for future investigation within the LFG-PFM framework is how changes in the Portuguese pronoun system interact with changes in the verbal paradigm. For example, changes in the second person pronouns and first person plural pronouns may be associated with the loss of the distinctions in verbal person/number inflection associated with these pronouns, leading to syncretism. The six person/number distinctions in the paradigm inherited from Latin are substantially reduced in some Portuguese varieties (Dubert and Galves 2016, 426), with second person and first person plural forms being particularly likely to be lost as grammaticalised pronouns (which tend to take third person singular agreement) gain frequency. This phenomenon raises questions about directionality: are the verb paradigm distinctions lost as a result of the increased use of the grammaticalised pronouns, or are the grammaticalised pronouns preferred because of general pressures to reduce verb paradigm complexity? Aalberse and Stoop (2015, 193) argue that in the history of English and Dutch, the inherited second person pronoun (the 'T' form: *thou* and *du*) are lost due to a combination of circumstances:

(a) There need to be sociolinguistic circumstances that strongly reduce the use of T.
(b) The loss of T should make the paradigm of verbal inflection more economical.
(c) The circumstances should create a need for a more economical inflectional paradigm; these circumstances would typically involve language contact situations.

The precise interaction between changes in the pronominal system and reductions in complexity of inflectional paradigms in the history of Portuguese, particularly in Brazil, is a fascinating area which has already been the object of some attention (Zilles 2005; Aalberse and Stoop 2015; Lara Bermejo and Guilherme 2018). LFG-PFM provides an ideal theoretical architecture for a formal approach to such changes, allowing a fine-grained representation of the interactions between the many different morphological, syntactic and semantic changes involved in the developments. A more detailed account of the diachrony of the widespread morphosyntactic change, and the differences in the speed and extent to which it has occurred in different Portuguese varieties,

represents an area for future research making use of the LFG-PFM integration of morphological, syntactic and semantic information.

More broadly, such an investigation could fruitfully incorporate both sociolinguistic detail and information on pragmatic (or discourse-level) phenomena, particularly since pronoun use is closely related to politeness phenomena, which are well known to vary greatly across speaker groups (see for example Nwoye 1992). There would be substantial motivation for incorporating information relating to both these areas into the LFG-PFM architecture laid out here. Although extensive work exists in LFG on information structure within sentences (Dalrymple, Lowe, and Mycock 2019, chap. 10), there remains little work on discourse structure, that is, regarding relationships across sentences or utterances. Similarly, there is scope to include within the architecture further detail about sociolinguistic implications, such as relating to politeness phenomena, but this has had little attention in previous LFG work.

Having seen some further questions raised by this chapter, and areas for future research, I now move on to summarise the conclusions drawn from my analysis of the Portuguese grammaticalised pronouns.

6.6 Conclusions

This chapter has demonstrated the usefulness of a formal LFG-PFM approach in accounting for the development of the grammaticalised pronoun forms (first person plural and second person) in Portuguese. I have shown that combining the modular approach of LFG with PFM insights into morphology allows the development of fine-grained representations of features at different levels. Changes in grammatical features, which often result in mismatches between different levels of the grammatical system, are a crucial part of the grammaticalisation process and must be included in any formal representation of diachronic change. The LFG-PGM approach is particularly well-suited to accounting for these mismatches.

Using the Portuguese pronouns as a case study, I have shown that an adequate theory must be able to capture mismatches between syntactic and semantic properties, and between different aspects of syntactic structure. Formalising these mismatches in the LFG framework, using functional structure and semantic structure, allows a clear representation of the key differences, which might go unnoticed in a less formal or less detailed account.

Grammaticalisation of a form from a canonical noun phrase to a pronoun also involves other changes which can be captured effectively in the LFG analysis, as we have seen: there can be changes in the value of the PRED feature, from a full semantic value to an anaphoric reference indicated by 'PRO', and changes in the value of PRONTYPE within the trajectory of a single form. These types of changes are a crucial part of grammaticalisation.

These changes go alongside changes in grammatical features undergone by these forms. For the pronouns studied here, we see that person, number, gender and case are important. I have introduced two adaptations to standard LFG representations to capture the complexities of the Portuguese data. These are including the PERSON feature in the CONCORD feature bundle at the level of f-structure, and including information on the person, number and gender of the referent in the semantic structure representation of the pronoun. These innovations allow an effective representation of agreement mismatches across different parts of the grammar. The first develops existing ideas on the representation of these mismatches to include previously unexplained patterns in the grammatical feature of person, while the second is crucial for explaining the progression of mismatches and correspondences between semantics and syntax.

This work highlights the importance of considering patterns which affect only a small number of items in the lexicon: in this case, I identify the importance of the mismatch behaviour associated with grammaticalised pronouns which have developed in the last several hundred years of the history of Portuguese. Taken separately, each of these pronouns presents a small exception to the normal patterns of grammatical features and agreement. Investigating these exceptions in more

detail provides a fascinating insight into the possible changes of grammatical features over time through the process of grammaticalisation. Comparing different varieties of Portuguese gives an insight into the types of changes that we see, and also clarifies that the feature changes associated with grammaticalisation may stop and become 'frozen' at any stage, leaving stable mismatches as a component of the grammatical system. Innovations are sometimes required to capture this kind of behaviour in a formal analysis, as we have seen particularly clearly for the person feature.

More generally, this chapter has demonstrated the wider possibilities of using LFG and PFM together to account for historical change. The flexibility of the LFG architecture, with its independent but mutually constraining modules, allows the creation of detailed models that can capture the state of the grammar at different points in time. We can represent long and complex processes of diachronic change as a series of incremental differences between grammars. The independence of each module is crucial: semantics, functional structure, constituent structure and morphology can all change without necessarily affecting the other modules at a single point in time.

The grammaticalised pronouns presented in this chapter represent a very different case study from the double past participles considered in Chapter 4 and Chapter 5, with mismatches occurring at the levels of functional and semantic structure, rather than within the morphology, but both phenomena can be analysed effectively using the LFG-PFM approach. In the final chapter of my thesis, I bring together my conclusions to present a wider perspective on the possibilities of this LFG-PFM approach to analysing synchronic mismatches and diachronic change.

Chapter 7 Conclusions from the thesis

In the introduction, I set out four research questions to be answered. These were:

1. How exactly does combining LFG with PFM work?
2. How can diachronic phenomena, including linguistic divergence, be analysed using LFG-PFM?
3. What new insights does the LFG-PFM analysis provide into the double past participle phenomenon and grammaticalised pronouns in Portuguese?
4. What extensions to the theory are needed to account for the Portuguese case studies?

In this chapter, I bring together my responses to these challenges. In section 7.1 I summarise the key contributions of the thesis, starting with more general theoretical contributions and then looking at aspects relating more specifically to Portuguese. Then in section 7.2 I suggest directions for future research building on the thesis.

7.1 Key contributions of the thesis

I first consider contributions made to LFG-PFM theory. In this thesis, I have applied a theoretical framework in which PFM acts as the realisational morphological component within the LFG architecture. PFM provides the realisation relation $\mathscr{R}$, giving the set of morphological entries required to enable the formation of a complete and correct set of lexical entries for the language. In each case, an inflected form is associated with its base lexeme and its morphosyntactic features. I have shown that a version of PFM based on Stump (2001) and extended by Taylor (2012) can be straightforwardly integrated into the LFG architecture, with less redundancy than an integration of PFM2 (Stump 2016).

This combination of LFG and PFM allows for syntactic and autonomously morphological phenomena, as well as phenomena that involve both syntax and morphology together, to be

treated together in a single theoretical architecture. This is particularly productive for analysing mismatches across syntax and morphology.

The separate levels of analysis, both within and across modules, mean the development of mismatches can be captured in detail. For example, we have seen mismatches within functional structure (between INDEX and CONCORD), and mismatches between functional structure and semantic structure. In each case, the existence of these separate levels of analysis has meant that even rare or apparently inconsequential mismatches can be given an adequate representation, allowing a more comprehensive account of the grammar.

This is further evidence in favour of the LFG convention of separating out different levels of analysis, to avoid conflating different aspects of the data. On the other hand, I follow recent LFG work which rejects a separate level of m-structure (Dalrymple 2015; Dalrymple, Lowe, and Mycock 2019, chap. 12), instead locating morphological information in the lexicon and using PFM to fulfil the role of the realisational morphological relation.

As well as demonstrating the usefulness of the LFG-PFM theory for synchronic analysis, I have shown that it allows fine-grained analysis of morphosyntactic theory in diachrony. Applying this theoretical perspective allows different aspects of diachronic change to be considered separately, making clear the incremental differences between historical varieties that add up to divergence over time and across space. By applying an LFG-PFM analysis to the separate varieties, we can see the intermediate stages in grammatical change, allowing new insights into the types of change that are more or less frequent, or the relationships between likelier and less likely types of change.

Turning more specifically towards the application of the theories to Portuguese, I have demonstrated the effectiveness of PFM in accounting for the morphology of Portuguese, including the double past participle phenomenon. Although there have been partial accounts of Portuguese morphology in PFM, these have not included an analysis of the double past participle phenomenon. My application of PFM to the double past participles, synchronically in Chapter 4

and diachronically in Chapter 5, has demonstrated that such an account can capture the full complexities of the phenomenon, including its arbitrary nature (lexical specification), and the changing relationship between regular/irregular forms and active/passive verbal contexts.

Similarly, in Chapter 6, I have shown that an analysis using LFG f-structures and s-structures, with some adaptations, is effective in accounting for person and number mismatches in Portuguese grammaticalised pronouns. These mismatches have not been considered before in LFG, and person mismatches more generally have been understudied in LFG work (and received relatively little attention from other perspectives).

As well as these theoretical contributions, this thesis has also made contributions to the study of Portuguese morphosyntax.

I have presented new insights into the double past participle phenomenon, particularly in terms of the types of diachronic change seen in its development across Portuguese. This has been facilitated by the rigorous application of PFM for analysing the forms' morphology. For example, I have shown that irregular stem roots tend to be lost from a verb's lexemic entry, rather than added. The one example of a novel irregular stem root was acquired through suppletion. In addition, my analysis has shown that inherited morphological irregularities such as past participle forms in /s/ or /t/ are very persistent, having come into Portuguese from Latin and still being maintained in the modern language. On the other hand, there is no evidence of similar irregularities in past participle forms being innovated. Innovations tend to build on existing stems, as we saw in the expansion of the truncated irregular past participle form across a larger number of verbs in the Brazilian innovative past participles. Since PFM allows an explicit account of verbs' arbitrary morphological patterns, including stems and inflection class membership, which are also represented in the LFG lexemic entry, LFG-PFM is particularly well-suited to capturing these facts.

I have given a new, clear presentation of evidence about mismatches of various types in varieties of Portuguese. In particular, I have brought together data on mismatches found across 2SG, 2PL and 1PL grammaticalised pronouns, analysing their grammatical features as marked on particular targets. This has extended our knowledge of what kind of mismatches exist in Portuguese, pointing to the possibility of future work investigating the typology of mismatches, as discussed below.

My work has also provided a new analysis of the Portuguese grammaticalised pronouns. We see a general move from syntactic to semantic agreement as the grammaticalisation process occurs, as expected. However, it is clear from the data across Portuguese varieties that these agreement patterns can become frozen at any time. More conservative varieties like standard European Portuguese are likely to retain syntactic agreement, particularly in person and number. I also suggested that due to independent changes in the inflectional paradigms (loss of person/number markings leading to syncretism), some changes that might be predicted, such as a move towards second person plural verb agreement with *vocês*, may never occur. This points to the interrelated nature of many morphological and syntactic changes, which can be represented more fully in a combined LFG-PFM framework than in a theoretical analysis that captures only syntax or only morphology.

Some extensions were required to standard LFG and PFM to capture fully the case studies analysed here. For the double past participles, I built on Taylor's (2012) adaptation of standard PFM, which includes rules for complex stem formation. This allowed me to provide newly comprehensive rules for Portuguese verbal inflectional morphology, including new definitions of inflection classes to capture verbs' behaviour in relation to the past participles. This analysis allows full and correct lexical entries to be given for verbs across varieties of Portuguese, whether they have a regular, an irregular, or both a regular and an irregular past participle. In accounting for the grammaticalised pronouns, I showed that some innovations in LFG are required to represent the full range of mismatch behaviour. These involved including PERSON as a feature within the CONCORD (as well

as INDEX) at f-structure, and including information on the person, number and gender of the referent in the semantic structure representation of the pronoun. This allows all three grammatical features to be represented at the three separate levels of CONCORD, INDEX and semantics, facilitating a fine-grained representation of mismatches between the levels.

A further contribution of this thesis relates to the types of data considered. The phenomena studied here are relatively marginal within the grammar of Portuguese, affecting small numbers of lexical items. However, in each case, we have seen that accounting for the wide range of behaviour requires some adjustments to the theoretical framework. This points to the value of studying apparently minor patterns in the grammar. Doing so gives us new insights into distinctions that could be easily overlooked when considering only 'well-behaved' grammatical phenomena. This expands our knowledge of what types of grammatical mismatch are possible, and what our analyses must account for.

7.2 Directions for future research

This section suggests directions for future research building on the work carried out in this thesis, from more direct extensions to more wide-ranging or speculative projects.

One question raised by this work is how the analysis of the phenomena considered here can be developed to account for other varieties of Portuguese, such as African Portuguese varieties. This would involve investigating the patterns of mismatch or inflectional morphology seen in those varieties, and adapting the LFG-PFM rules. Such an analysis could shed more light on the types of grammatical divergence seen within Portuguese, while also clarifying any further necessary developments to the theoretical framework.

Along similar lines, it would be illuminating to make an LFG-PFM analysis of the patterns seen in other Romance languages that are similar to the Portuguese double past participle, such as the

overabundant past participles found in some Italian varieties. This would clarify similarities and differences between Romance varieties.

In Chapter 6, we saw a range of mismatches across the noun phrase in Portuguese. This raises the question of what other types of mismatches are found in Portuguese noun phrases, and how these should be analysed. Working with Wechsler and Zlatic's schema, repeated in (279), the grammaticalised pronouns I have considered showed mismatches between CONCORD and INDEX, and between INDEX and SEMANTICS. As mentioned in section 6.5.3, there are some possible mismatches between DECLENSION and CONCORD in Portuguese, which should be investigated further.

(279) DECLENSION | CONCORD ↔ INDEX ↔ SEMANTICS

Extending this idea more generally, it would be fruitful to study the full typology of mismatches found across Portuguese, in the verb phrase and more generally, as well as in the noun phrase.

I have suggested a possible analysis for mismatches in the PERSON feature in Portuguese, treating it as a feature in both INDEX and CONCORD. This raises the question of whether person mismatches in other languages, such as imposters, can be treated in the same manner, or whether an alternative analysis (more in line with original proposals relating to the distinct historical origin of INDEX versus CONCORD features) might be more appropriate.

This research also opens up broader questions about applying a formal theoretical perspective to the diachrony of morphology. Applying LFG-PFM to a wider study of inflectional change across Portuguese varieties, and indeed pidgins and creoles related to Portuguese, could result in new insights into the broader types of morphological change that are seen. For example, we might find tendencies across the types of PFM rules that are lost or retained, or whose application is reduced or extended: perhaps there is a relationship between the narrowness of rules and their tendency towards loss or retention.

One particular area of interest within Portuguese morphology is the rise of syncretism shown in the generalisation of third person singular verb agreement across the paradigm, with the loss of distinct inflection patterns for other person and number combinations. We saw in Chapter 6 that this leads to the second person singular pronoun *tu* being used with third person singular verb forms. A future programme of research could use LFG-PFM to investigate the diachronic interactions between syntax and morphology in a broader sense, for example looking at a series of interrelated changes taking place in Brazilian Portuguese: the rise of grammaticalised pronouns, the loss of person/number distinctions in the verb paradigm, and a reduction in null subjects in favour of overt subjects. All of these phenomena have their own complexities but also interact in various ways, with striking differences between varieties. This presents an ideal testing ground for the LFG-PFM theory, providing opportunities to develop the framework's ability to account for interrelated diachronic changes that affect several parts of the grammar.

Another interesting area for research arising from applying PFM to a diachronic perspective is the possibility of considering changes that occur across paradigms, such as suppletion, the merging of two verb paradigms into one, and analogy across lexemes. Stump (2001, sec. 7.5) discusses 'translexemic syncretisms', in which one verb's paradigm is partially syncretic with another's, from a synchronic perspective, but does not consider their diachrony. It would be interesting to investigate how such patterns arise, as part of a more general investigation of cross-paradigm morphological change, using PFM. Integrating Spencer's (2013) GPFM account of relationships across lexemes might also be useful for this purpose. Such an account could incorporate information from different modules of the grammar including semantics, inflection class and morphological form, as well as functional information such as frequency.

There are also further possibilities for applying LFG-PFM to aspects of language not considered here. For example, building on my analysis of number and gender agreement of the past participle in section 4.4.7 and the morphological properties of nouns and pronouns in section 6.3.2, it would

be worthwhile to develop a full PFM account of nominal morphology in Portuguese, including an analysis of the inflection classes of nouns, a treatment of diminutives, and an account of gender in adjectives and nouns. In the latter cases, a GPFM account of the lexical relatedness of forms with different gender properties and of diminutives might give useful new insights. A full formal account of nominal morphology could feed into a more comprehensive account of Romance morphology across all different lexical categories.

Beyond syntax and morphology, there is a rich range of possibilities to engage with other aspects of language and language history from the LFG-PFM perspective. One possible programme of research would investigate the links between language contact and morphosyntactic change across Portuguese varieties, including Brazilian Portuguese, African Portuguese and Portuguese-based creoles. In this context, having countable measures, such as number of rules in particular PFM rule blocks, provides a clear quantitative advantage over more qualitative approaches to historical morphosyntactic change. This points towards the wider possibilities for the LFG-PFM framework, which facilitates a rigorous approach to both synchronic and diachronic phenomena.

Appendices

Appendix I Verbal paradigms: simplex forms

Paradigms are given in the order: 1SG, 2SG, 3SG, 1PL, 2PL, 3PL

Levar 'carry' (1st conjugation, regular)

Indicative

Present		Preterite		Imperfect	
eu	levo	eu	levei	eu	levava
tu	levas	tu	levaste	tu	levavas
ele/ela	leva	ele/ela	levou	ele/ela	levava
nós	levamos	nós	levámos	nós	levávamos
vós	levais	vós	levastes	vós	leváveis
eles/elas	levam	eles/elas	levaram	eles/elas	levavam

Pluperfect		Future		Conditional	
eu	levara	eu	levarei	eu	levaria
tu	levaras	tu	levarás	tu	levarias
ele/ela	levara	ele/ela	levará	ele/ela	levaria
nós	leváramos	nós	levaremos	nós	levaríamos
vós	leváreis	vós	levareis	vós	levaríeis
eles/elas	levaram	eles/elas	levarão	eles/elas	levariam

Subjunctive

Present		Imperfect		Future	
eu	leve	eu	levasse	eu	levar
tu	leves	tu	levasses	tu	levares
ele/ela	leve	ele/ela	levasse	ele/ela	levar
nós	levemos	nós	levássemos	nós	levarmos
vós	leveis	vós	levásseis	vós	levardes
eles/elas	levem	eles/elas	levassem	eles/elas	levarem

Non-finite forms

Personal infinitive		Infinitive	Gerund	Past participle
eu	levar	levar	levando	levado
tu	levares			
ele/ela	levar			
nós	levarmos			
vós	levardes			
eles/elas	levarem			

***Beber* 'drink' (2nd conjugation, regular)**

Indicative

Present		Preterite		Imperfect	
eu	bebo	eu	bebi	eu	bebia
tu	bebes	tu	bebeste	tu	bebias
ele/ela	bebe	ele/ela	bebeu	ele/ela	bebia
nós	bebemos	nós	bebemos	nós	bebíamos
vós	bebeis	vós	bebestes	vós	bebíeis
eles/elas	bebem	eles/elas	beberam	eles/elas	bebiam

Pluperfect		Future		Conditional	
eu	bebera	eu	beberei	eu	beberia
tu	beberas	tu	beberás	tu	beberias
ele/ela	bebera	ele/ela	beberá	ele/ela	beberia
nós	bebêramos	nós	beberemos	nós	beberíamos
vós	bebêreis	vós	bebereis	vós	beberíeis
eles/elas	beberam	eles/elas	beberão	eles/elas	beberiam

Subjunctive

Present		Imperfect		Future	
eu	beba	eu	bebesse	eu	beber
tu	bebas	tu	bebesses	tu	beberes
ele/ela	beba	ele/ela	bebesse	ele/ela	beber
nós	bebamos	nós	bebêssemos	nós	bebermos
vós	bebais	vós	bebêsseis	vós	beberdes
eles/elas	bebam	eles/elas	bebessem	eles/elas	beberem

Non-finite forms

Personal infinitive		Infinitive	Gerund	Past participle
eu	beber	beber	bebendo	bebido
tu	beberes			
ele/ela	beber			
nós	bebermos			
vós	beberdes			
eles/elas	beberem			

***Partir* 'break' (3rd conjugation, regular)**

Indicative

Present		Preterite		Imperfect	
eu	parto	eu	parti	eu	partia
tu	partes	tu	partiste	tu	partias
ele/ela	parte	ele/ela	partiu	ele/ela	partia
nós	partimos	nós	partimos	nós	partíamos
vós	partis	vós	partistes	vós	partíeis
eles/elas	partem	eles/elas	partiram	eles/elas	partiam

Pluperfect		Future		Conditional	
eu	partira	eu	partirei	eu	partiria
tu	partiras	tu	partirás	tu	partirias
ele/ela	partira	ele/ela	partirá	ele/ela	partiria
nós	partíramos	nós	partiremos	nós	partiríamos
vós	partíreis	vós	partireis	vós	partiríeis
eles/elas	partiram	eles/elas	partirão	eles/elas	partiriam

Subjunctive

Present		Imperfect		Future	
eu	parta	eu	partisse	eu	partir
tu	partas	tu	partisses	tu	partires
ele/ela	parta	ele/ela	partisse	ele/ela	partir
nós	partamos	nós	partíssemos	nós	partirmos
vós	partais	vós	partísseis	vós	partirdes
eles/elas	partam	eles/elas	partissem	eles/elas	partirem

Non-finite forms

Personal infinitive		Infinitive	Gerund	Past participle
eu	partir	partir	partindo	partido
tu	partires			
ele/ela	partir			
nós	partirmos			
vós	partirdes			
eles/elas	partirem			

***Ser* 'be' (irregular)**

Indicative

Present		Preterite		Imperfect	
eu	sou	eu	fui	eu	era
tu	és	tu	foste	tu	eras
ele/ela	é	ele/ela	foi	ele/ela	era
nós	somos	nós	fomos	nós	éramos
vós	sois	vós	fostes	vós	éreis
eles/elas	são	eles/elas	foram	eles/elas	eram

Pluperfect		Future		Conditional	
eu	fora	eu	serei	eu	seria
tu	foras	tu	serás	tu	serias
ele/ela	fora	ele/ela	será	ele/ela	seria
nós	fôramos	nós	seremos	nós	seríamos
vós	fôreis	vós	sereis	vós	seríeis
eles/elas	foram	eles/elas	serão	eles/elas	seriam

Subjunctive

Present		Imperfect		Future	
eu	seja	eu	fosse	eu	for
tu	sejas	tu	fosses	tu	fores
ele/ela	seja	ele/ela	fosse	ele/ela	for
nós	sejamos	nós	fôssemos	nós	formos
vós	sejais	vós	fôsseis	vós	fordes
eles/elas	sejam	eles/elas	fossem	eles/elas	forem

Non-finite forms

Personal infinitive		Infinitive	Gerund	Past participle
eu	ser	ser	sendo	sido
tu	seres			
ele/ela	ser			
nós	sermos			
vós	serdes			
eles/elas	serem			

***Ter* 'have' (irregular)**

Indicative

Present		Preterite		Imperfect	
eu	tenho	eu	tive	eu	tinha
tu	tens	tu	tiveste	tu	tinhas
ele/ela	tem	ele/ela	teve	ele/ela	tinha
nós	temos	nós	tivemos	nós	tínhamos
vós	tendes	vós	tivestes	vós	tínheis
eles/elas	têm	eles/elas	tiveram	eles/elas	tinham

Pluperfect		Future		Conditional	
eu	tivera	eu	terei	eu	teria
tu	tiveras	tu	terás	tu	terias
ele/ela	tivera	ele/ela	terá	ele/ela	teria
nós	tivéramos	nós	teremos	nós	teríamos
vós	tivéreis	vós	tereis	vós	teríeis
eles/elas	tiveram	eles/elas	terão	eles/elas	teriam

Subjunctive

Present		Imperfect		Future	
eu	tenha	eu	tivesse	eu	tiver
tu	tenhas	tu	tivesses	tu	tiveres
ele/ela	tenha	ele/ela	tivesse	ele/ela	tiver
nós	tenhamos	nós	tivéssemos	nós	tivermos
vós	tenhais	vós	tivésseis	vós	tiverdes
eles/elas	tenham	eles/elas	tivessem	eles/elas	tiverem

Non-finite forms

Personal infinitive		Infinitive	Gerund	Past participle
eu	ter	ter	tendo	tido
tu	teres			
ele/ela	ter			
nós	termos			
vós	terdes			
eles/elas	terem			

Appendix II Periphrastic or compound verb forms

ACTIVE compound forms			**Tense of auxiliary**
compound past	*tem comido*	'he has been eating'	Present indicative
compound pluperfect	*tinha comido*	'he had eaten'	Imperfect indicative
future perfect	*terá comido*	'he will have eaten'	Future indicative
	**teve comido*[145]		Past
past conditional	*teria comido*	'he would have eaten'	Conditional
PASSIVE compound forms			**Tense of auxiliary**
present passive	*é comido*	'it is eaten'	Present indicative
imperfect passive	*era comido*	'it was eaten' (state or habitual)	Imperfect indicative
future passive	*será comido*	'it will be eaten'	Future indicative
past passive	*foi comido*	'it was eaten' (punctual)	Past
conditional passive	*seria comido*	'it would be eaten'	Conditional

Table 50 Periphrastic forms with a single auxiliary verb

[145] This form does not exist. For brief discussion of this form and a comparison with other Romance varieties, see Oliveira (2013, 528) and Laca (2004, 438).

Compound form			**Tense of auxiliary**
perfect passive	*tem sido comido*	'it has been eaten'	Present indicative
pluperfect passive	*tinha sido comido*	'it had been eaten'	Imperfect indicative
future perfect passive	*terá sido comido*	'it will have been eaten'	Future indicative
conditional perfect passive	*teria sido comido*	'it would have been eaten'	Conditional

Table 51 Periphrastic forms with two auxiliary verbs

All the forms given above are in the indicative mood. Some of these forms can be made subjunctive by the use of the corresponding subjunctive tensed auxiliary, where the tensed auxiliary is in a tense that has a corresponding subjunctive, as below. This is an example of grammatical compositionality.

ACTIVE compound subjunctive forms			**Tense of auxiliary**
perfect subjunctive	*tenha comido*	'he has eaten'	present subjunctive
pluperfect subjunctive	*tivesse comido*	'he would have eaten'	imperfect subjunctive
future perfect subjunctive	*tiver comido*	'he would eat' (rare)	future subjunctive
PASSIVE compound subjunctive forms			**Tense of auxiliary**
present passive subjunctive	*seja comido*	'it would be eaten'	present subjunctive
imperfect passive subjunctive	*fosse comido*	'it would have been eaten'	imperfect subjunctive
future passive subjunctive	*for comido*	'it would be eaten' (rare)	future subjunctive

Table 52 Periphrastic forms with subjunctive auxiliary verb

perfect passive subjunctive	*tenha sido comido*	'it has been eaten'
pluperfect passive subjunctive	*tivesse sido comido*	~ 'it would have been eaten'
future passive subjunctive	*tiver sido comido*	~ 'it will have been eaten'

Table 53 Periphrastic forms with subjunctive of *ter* 'have' auxiliary and *ser* 'be' auxiliary

Appendix III PFM rules and lexical data for Portuguese

Rules are adapted from Taylor (2012) and renumbered; my adaptations to account for irregular and double past participles are highlighted in green.

Verb classes

101. **Pt** = the set of all Portuguese lexemes

102. **C1** ∩ **Pt** = the set of all Portuguese /ar/ verbs

103. **C2** ∩ **Pt** = the set of all Portuguese /er/ verbs

104. **C3** ∩ **Pt** = the set of all Portuguese /ir/ verbs

105. (**C1** ∪ **C2** ∪ **C3**) ∩ **Pt** = **V** ∩ **Pt**

106. {CABER, DIZER, ESTAR, FAZER, HAVER, IR, PODER, PÔR, QUERER, SABER, SER, TER, TRAZER, VIR, ...} = **PYTA** ⊂ **V** ∩ **Pt**

107. {CABER, CRER, DIZER, ESTAR, FAZER, HAVER, LER, MEDIR, OUVIR, PEDIR, PERDER, PODER, PÔR, QUERER, RIR, SABER, SER, TER, TRAZER, VER, VIR, ...} = **LPattern** ⊂ **V** ∩ **Pt**

108. {ENTREGAR, GANHAR, GASTAR, LIMPAR, MATAR, PAGAR …} = **TruncPTCP** ⊂ **V** ∩ **Pt**

109. {ACENDER, EXPRIMIR, PRENDER, SUSPENDER …} = **S-IrregPTCP** ⊂ **V** ∩ **Pt**

110. {ABRIR, COBRIR, DIZER, ELEGER, ESCREVER, FAZER, MORRER, PÔR, VER …} = **T-IrregPTCP** ⊂ **V** ∩ **Pt**

111. **S-IrregPTCP** ∪ **T-IrregPTCP** ∪ **TruncPTCP** = **IrregPTCP** ⊂ **V** ∩ **Pt**

112. {ACENDER, PRENDER, EXPRIMIR …} = **DoublePTCP** ⊂ **V** ∩ **Pt**

113. {SER, IR, RIR, LER, VER} = **pplImp** ⊂ **V** ∩ **Pt**

114. {TER, IR, VER, PÔR} = **gerImp** ⊂ **V** ∩ **Pt**

115. **IrrImp** = **pplImp** ∪ **gerImp**

116. {IR, SER, DAR} = **1sgPI** ⊂ **V** ∩ **Pt**

Stem formation rules

- Block 0 stem formation rules

117. BSR_0,{g,h,k,n},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[s-rootg], ι>

118. BSR_0,{h},**PYTA** ∩ **Pt**(<χ, ι>) =$_{def}$ <[s-rooth], ι>

119. BSR_0,{k},**LPattern** ∩ **Pt**(<χ, ι>) =$_{def}$ <[s-rootk], ι>

120. BSR_0,{l},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[s-rootl], ι>

121. BSR_0,{a,b,d,e},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[stemg[7]], ι>

122. BSR_0,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[stemg[6]], ι>

123. BSR_0,{ɛ},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[steml[9]], ι>

124. BSR_0,{f},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[stemh[4]], ι>

- Block 1 stem formation rules

125. SR_1,{a,b,f},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χ_i/r/], ι> for all i

126. SR_1,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/d/], ι>for all i

127. SR_1,{ɛ},**T-IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/t/], ι> for all i

128. SR_1,{ɛ},**S-IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/s/], ι> for all i

129. SR_1,{ɛ},**TruncPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi], ι>for all i

130. SR_1,{e},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χ_i/nd/], ι> for all i

131. SR_1,{d},**pplImp**(<χ, ι>) =$_{def}$ <γ, ι>, where Nar1,**V** ∩ **Pt**(<χ, c>) = <γ, c>

132. SR_1,{d},**gerImp**(<χ, ι>) =$_{def}$ <γ, ι>, where Nar1,**V** ∩ **Pt**(<χ, e>) = <γ, e>

Stem selection rules

- Basic stem selection rules

133. SSR_0,←{}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemg [1]], σ>

134. where τ =← {tns: pres, md: subj} →,

SSR_0,τ,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemk [2]], σ>

↕

SSR_0,τ/{agr: {per: 1, num: sg},md: ind},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemk [2]], σ>

135. where τ =← {tns: past 2, md: ind} →,

SSR_0,τ,**PYTA** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemh [4]], σ>

↕

SSR_0,τ/{tns: plup,md: ind},**PYTA** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemh [4]], σ>

↕

SSR_0,τ/{tns: past 1,md: subj},**PYTA** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemh [4]], σ>

136. where τ =← {tns: pres, md: ind} →,

τ ≠← {agr: {per: 2, num: pl}} →

and ϱ = {md: subj}

SSR_0,τ,{IR}∩ **Pt**(<χ, σ>) =$_{def}$ <ψ, σ>, where Nar0,{IR}∩ **Pt**(<χ, σ/ϱ>) = <ψ, σ/ϱ>

137. SSR_0,←{tns: past 1,md: ind}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemn [1]], σ>

138. SSR_0,←{tns: past 1,md: ind}→,(**C2∪C3**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemn [3]], σ>

139. SSR_0,←{tns: past 1,md: ind}→,{SER}∩ **Pt**(<χ, σ>) =$_{def}$ <[stemn [10]], σ>

- Complex stem selection rules

140. SSR_0,←{vform: inf}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stema [10]], σ>

141. SSR_0,←{tns: fut,md: ind}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemb [10]], σ>

142. SSR_0,←{tns: cond}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemb [3]], σ>

143. SSR_0,{vform: ppl},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemc [10]], σ>

144. SSR_0,{vform: ppl}, (**IrregPTCP**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemɛ [8]], σ>

145. SSR_0,{vform: ppl, voice: active}, (**DoublePTCP**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemc [10]], σ>

146. SSR_0,{agr: {per: 2, num: pl},md: imp},**IrrImp** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemd [5]], σ>

147. SSR_0,{vform: ger},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[steme [12]], σ>

148. where α = 1,2,3, β = sg,pl,

SSR_0,{agr: {per: α, num: β},tns: fut,md: subj},**PYTA** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemf [10]], σ>

Whole word syncretism rules

149. where n = 0,1,2,

τ = ← {agr: {per: 1, num: sg}} but not of

{tns: pres, md: ind}, {tns: past 2, md: ind},

{tns: plup, md: ind}or {tns: fut, md: ind} →

and ϱ = {agr: {per: 3}}

RR_n,τ, **V** ∩ **Pt**(<χ, σ>) =$_{def}$ <ψ, σ>, where Nar_n, **V** ∩ **Pt**(<χ, σ/ϱ>) = <ψ, σ/ϱ>

150. where n = 0,1,2,

τ = {agr: {per: 2, num: sg}, md: imp}

and ϱ = {agr: {per: 3}, tns: pres, md: ind}

RR_n,τ, **V** ∩ **Pt**(<χ, σ>) =$_{def}$ <ψ, σ>, where Nar_n, **V** ∩ **Pt**(<χ, σ/ϱ>) = <ψ, σ/ϱ>

151. where n = 0,1,2,

τ = {agr: {per: 3, num: pl}, tns: past 2, md: ind}

and ϱ = {tns: plup}

RR_n,τ, **V** ∩ **Pt** (<χ, σ>) =$_{def}$ <ψ, σ>, where Nar_n,**V** ∩ **Pt**(<χ, σ/ϱ>) = <ψ, σ/ϱ>

152. where n = 0,1,2,

τ = {vform: ppl}

and ϱ = {vform: ger}

RR_n,τ,{vir} ∩ **Pt**(<χ, σ>) =$_{def}$ <ψ, σ>, where Nar_n,{vir} ∩ **Pt**(<χ, σ/ϱ>) = <ψ, σ/ϱ>

153. where n = 0,1,2,

τ = {agr: {per: α, num: β}, tns: fut, md: subj}

and ϱ = {agr: {per: α, num: β}, vform: inf}

where α and β are any permitted values

$RR_{n},\tau,\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <\psi, \sigma>$, where $Nar_{n},\mathbf{V} \cap \mathbf{Pt}(<\chi, \varrho>) = <\psi, \varrho>$

Realisation rules

- Block 1 realisation rules

154. $RR_{1},\leftarrow\{tns: pres,md: ind\}\rightarrow,\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <[\chi_{i}/o/], \sigma>$ for all i

155. where τ = {agr: {per: 1, num: pl}, tns: pres, md: ind}

and ϱ = {agr: {per: 1, num: sg}}

$RR_{1},\tau,\{SER\} \cap Pt(<\chi, \sigma>) =_{def} <\psi, \sigma>$, where Nar1, $\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma/\varrho>) = <\psi, \sigma/\varrho>$

↕

$RR_{1},\tau/\{agr: \{per: 2\}\},\{SER\} \cap Pt(<\chi, \sigma>) =_{def} <\psi, \sigma>$, where

Nar1,$\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma/\varrho>) = <\psi, \sigma/\varrho>$

156. $RR_{1},\leftarrow\{tns: past 1,md: ind\}\rightarrow, \mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <[\chi_{i}/va/], \sigma>$ for all i

157. where $\tau = \leftarrow$ {tns: past 1, md: ind} $\rightarrow$

and ϱ = {tns: cond},

$RR_{1},\tau,(\mathbf{C2}\cup\mathbf{C3}) \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <\psi, \sigma>$, where Nar1,$\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma/\varrho>) = <\psi, \sigma/\varrho>$

158. $RR_{1},\leftarrow\{tns: cond\}\rightarrow,\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <[\chi_{i}/a/], \sigma>$ for all i

159. $RR_{1},\leftarrow\{tns: past 1,md: subj\}\rightarrow,\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <[\chi_{i}/se/], \sigma>$ for all i

160. $RR_{1},\{agr: \{per: 1, num: sg\},tns: past 2,md: ind\},\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <[\chi_{i}/i/], \sigma>$ for all i

161. $RR_{1},\leftarrow\{agr: \{per: 2\},tns: past 2,md: ind\}\rightarrow,\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <[\chi_{i}/ste/], \sigma>$ for all i

162. $RR_{1},\{agr: \{per: 3, num: sg\},tns: past 2,md: ind\},(\mathbf{C1}\backslash\mathbf{PYTA}) \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <[\chi_{i}/o/], \sigma>$ for all i

163. $RR_{1},\{agr: \{per: 3, num: sg\},tns: past 2,md: ind\},(\mathbf{V}\backslash\mathbf{PYTA}) \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <[\chi_{i}/N/], \sigma>$ for all i

164. $RR_{1},\leftarrow\{tns: plup\}\rightarrow,\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <[\chi_{i}/ra/], \sigma>$ for all i

165. $RR_{1},\leftarrow\{agr: \{per: 1\},tns: fut,md: ind\}\rightarrow,\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <[\chi_{i}/é/], \sigma>$ for all i

166. where τ = {agr: {per: 2, num: pl}, tns: fut, md: ind}

and ϱ = {agr: {per: 1}}

RR_1,τ,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <ψ, σ>, where Nar1,**V** ∩ **Pt**(<χ, σ/ϱ>) = <ψ, σ/ϱ>

167. RR_1,←{agr: {per: 3},tns: fut,md: ind}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/á/], σ> for all i

168. where τ = {agr: {per: 2, num: sg}, tns: fut, md: ind}

and ϱ = {agr: {per: 3}}

RR_1,τ,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <ψ, σ>, where Nar1,**V** ∩ **Pt**(<χ, σ/ϱ>) = <ψ, σ/ϱ>

169. where α = 1,2,3,β = sg,pl,

τ = {agr: {per: α, num: β}, tns: pres, md: ind}

and ϱ = {tns: fut}

RR_1,τ,{HAVER} ∩ **Pt**(<χ, σ>) =$_{def}$ <ψ, σ>, where Nar1,**V** ∩ **Pt**(<χ, σ/ϱ>) = <ψ, σ/ϱ>

170. RR_1,{vform: ppl, agr: {gend: fem},**V** ∩ **Pt**(<χo, σ>) =$_{def}$ <[χ_i/a/], σ> for all i

- Block 2 realisation rules

171. RR_2,{agr: {per: 1, num: sg},tns: pres,md: ind},**1sgPI** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/u/], σ> for all i

172. RR_2,←{agr: {per: 2, num: sg}}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/s/], σ> for all i

173. RR_2,←{agr: {per: 2, num: pl},md: imp}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/i/], σ> for all i

174. RR_2,←{agr: {per: 1, num: pl}}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/mos/], σ> for all i

175. RR_2,←{agr: {per: 2, num: pl}}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/is/], σ> for all i

176. RR_2,{agr: {per: 2, num: pl},vform: inf},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/des/], σ> for all i

177. RR_2,←{agr: {per: 3, num: pl}}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/u/], σ> for all i

178. where n = 2,

τ is {agr: {per: 3, num: pl}, tns: pres, md: ind},

ϱ {agr: {per: 3, num: pl}, vform: inf}

$RR_{n,\tau,}$(**C2∪C3**) ∩ **Pt**(<χ, σ>) $=_{def}$ <[χ_i/i/], σ> for all i

↕

$RR_{n,\tau}$/{tns: past 1,md: subj},**V** ∩ **Pt**(<χ, σ>) $=_{def}$ <[χ_i/i/], σ> for all i

↕

$RR_{n,\varrho}$,**V** ∩ **Pt**(<χ, σ>) $=_{def}$ <[χ_i/i/], σ> for all i

179. RR_2,{vform: ppl, agr: {num: pl},**V** ∩ **Pt**(<χ, σ>) $=_{def}$ <[χ_i/s/], σ> for all i

Lexical Data

Two tables of lexical data are given for a selection of Portuguese verbs, both regular and irregular: the first shows the basic stem roots of each verb, and the second the theme vowel array.

In the tables of lexical data, verbs not given in Taylor's (2012) analysis are shown below the dotted line. I follow Taylor in using a simplified phonemic representation rather than standard orthography.

Basic stem roots

	English translation	**sr$_g$**	**sr$_h$**	**sr$_k$**	**sr$_l$**
CANTAR	sing	kant	—	—	—
COMER	eat	kom	—	—	—
PARTIR	break	part	—	—	—
CABER	fit	kab	kob	kaib	—
CRER	believe	kr	—	krei	—
DAR	give	d	—	—	—
DIZER	say	diz	dis	dig	di
ESTAR	be	est	estiv	esteZ	—
FAZER	do	faz	fiz	fas	fei
LER	read	l	—	lei	—
MEDIR	measure	med	—	mes	—
OUVIR	hear	ov	—	os, oj	—
PEDIR	ask	ped	—	pes	—
PERDER	lose	perd	—	perk	—
PODER	be able	pod	pud	pos	—
QUERER	want	ker	kis	keir	—
RIR	laugh	r	—	ri	—
SABER	know	sab	sob	saib	—
TRAZER	bring	traz	troS	trag	—
VALER	be worth	val	—	vaL	—
ABRIR	open	abr	—	—	aber
ACEITAR	accept	aseit	—	—	—
ACENDER	light, turn on	asend	—	—	ase
COBRIR	cover	kobr	—	—	kober
ELEGER	choose	eleZ	—	—	elei
ENTREGAR	hand over	entreg	—	—	—
ESCREVER	write	eskrev	—	—	eskri
EXPRIMIR	express	esprim	—	—	espres
GANHAR	earn	gaɲ	—	—	—
GASTAR	spend	gast	—	—	—
LIMPAR	clean	limp	—	—	—
MORRER	die	morr	—	—	mor
PAGAR	pay	pag	—	—	—
PÔR	put	p	—	poɲ	pos
PRENDER	arrest	prend	—	—	pre
SUSPENDER	suspend	suspend	—	—	suspen
VER	see	v	—	veZ	vis

Theme vowel array

	TV[1]	TV[2]	TV[3]	TV[4]	TV[5]	TV[6]	TV[7]	TV[8]
CANTAR	a	e	i	a	a	a	a	o
COMER	e	a	i	e	e	i	e	o
PARTIR	i	a	i	i	e	i	i	o
CABER	e	a	i	e	e	i	e	o
CRER	e	a	i	e	a	i	e	o
DIZER	e	a	i	e	a	i	∅	o
ESTAR	a	a	i	e	e	a	a	o
FAZER	e	a	i	e	a	i	∅	o
LER	e	a	i	e	e	i	e	o
MEDIR	i	a	i	i	i	i	i	o
OUVIR	i	a	i	i	i	i	i	o
PEDIR	i	a	i	i	i	i	i	o
PERDER	e	a	i	e	e	i	e	o
PODER	e	a	i	e	e	i	e	o
QUERER	e	a	i	e	e	i	e	o
RIR	e	a	i	i	i	i	e	o
SABER	e	a	i	e	i	i	e	o
TRAZER	e	a	i	e	e	i	∅	o
VALER	e	a	i	e	e	i	e	o
ABRIR	i	a	i	i	e	i	i	o
ACEITAR	a	r	i	a	a	a	a	e
ACENDER	e	a	i	e	e	i	e	o
COBRIR	i	a	i	i	e	i	i	o
ELEGER	e	a	i	e	e	i	e	o
ENTREGAR	a	e	i	a	a	a	a	e
ESCREVER	e	a	i	e	e	i	e	o
EXPRIMIR	i	a	i	i	e	i	i	o
GANHAR	a	r	i	a	a	a	a	o
GASTAR	a	a	i	a	a	a	a	o
LIMPAR	a	r	i	a	a	a	a	o
MORRER	e	a	i	e	e	i	e	o
PAGAR	a	r	i	a	a	a	a	o
PÔR	e	a	i	e	e	i	e	o
PRENDER	e	a	i	e	e	i	e	o
SUSPENDER	e	a	i	e	e	i	e	o
VER	e	a	i	a	e	i	e	o

	TV[9]	TV[10]
all verbs	∅	o

Appendix IV PFM rule blocks for Brazilian Portuguese past participles

The rules have not been adapted to bring them closer to Brazilian Portuguese in other respects, e.g. loss of second person plural verb forms which are not found in Brazilian Portuguese varieties.

Verb classes

101. **Pt** = the set of all Portuguese lexemes

102. **C1** ∩ **Pt** = the set of all Portuguese /ar/ verbs

103. **C2** ∩ **Pt** = the set of all Portuguese /er/ verbs

104. **C3** ∩ **Pt** = the set of all Portuguese /ir/ verbs

105. (**C1** ∪ **C2** ∪ **C3**) ∩ **Pt** = **V** ∩ **Pt**

106. {CABER, DIZER, ESTAR, FAZER, HAVER, IR, PODER, PÔR, QUERER, SABER, SER, TER, TRAZER, VIR, ...} = **PYTA** ⊂ **V** ∩ **Pt**

107. {CABER, CRER, DIZER, ESTAR, FAZER, HAVER, LER, MEDIR, OUVIR, PEDIR, PERDER, PODER, PÔR, QUERER, RIR, SABER, SER, TER, TRAZER, VER, VIR, ...} = **LPattern** ⊂ **V** ∩ **Pt**

108. {ENTREGAR, GANHAR, GASTAR, LIMPAR, MATAR, PAGAR …} = **TruncPTCP** ⊂ **V** ∩ **Pt**

For Brazilian Portuguese innovative past participles:

108b. {ACEITAR, CHEGAR, COMPRAR, ENTREGAR, FALAR, FICAR, GANHAR, GASTAR, LIMPAR, PAGAR, PEDIR, PEGAR, PERDER, TRAZER …} = **TruncPTCP** ⊂ **V** ∩ **Pt** *edited*

109. {ACENDER, EXPRIMIR, PRENDER, SUSPENDER …} = **S-IrregPTCP** ⊂ **V** ∩ **Pt**

110. {ABRIR, COBRIR, DIZER, ELEGER, ESCREVER, FAZER, MORRER, PÔR, VER …} = **T-IrregPTCP** ⊂ **V** ∩ **Pt**

111. **S-IrregPTCP ∪ T-IrregPTCP ∪ TruncPTCP = IrregPTCP ⊂ V ∩ Pt**

112. {ACENDER, PRENDER, EXPRIMIR …} = **DoublePTCP ⊂ V ∩ Pt**

113. {SER, IR, RIR, LER, VER} = **pplImp ⊂ V ∩ Pt**

114. {TER, IR, VER, PÔR} = **gerImp ⊂ V ∩ Pt**

115. **IrrImp = pplImp ∪ gerImp**

116. {IR, SER, DAR} = **1sgPI ⊂ V ∩ Pt**

For a variety with no active/passive distinction:

108b. {ACEITAR, CHEGAR, COMPRAR, ENTREGAR, FALAR, FICAR, GANHAR, GASTAR, LIMPAR, PAGAR, PEDIR, PEGAR, PERDER, TRAZER …} = **TruncPTCP ⊂ V ∩ Pt**

109b. {ACENDER, EXPRIMIR, PRENDER, SUSPENDER …} = **S-IrregPTCP ⊂ V ∩ Pt**

110b. {ABRIR, COBRIR, DIZER, ELEGER, ESCREVER, FAZER, MORRER, PÔR, VER …} = **T-IrregPTCP ⊂ V ∩ Pt**

111b. **S-IrregPTCP ∪ T-IrregPTCP ∪ TruncPTCP = IrregPTCP ⊂ V ∩ Pt**

~~112. {ACENDER, PRENDER, EXPRIMIR …} = **DoublePTCP ⊂ V ∩ Pt**~~

Stem formation rules

- Block 0 stem formation rules

117. BSR_0,{g,h,k,n},**V ∩ Pt**(<χ, ι>) =$_{def}$ <[s-rootg], ι>

118. BSR_0,{h},**PYTA ∩ Pt**(<χ, ι>) =$_{def}$ <[s-rooth], ι>

119. BSR_0,{k},**LPattern ∩ Pt**(<χ, ι>) =$_{def}$ <[s-rootk], ι>

120. BSR_0,{l},**V ∩ Pt**(<χ, ι>) =$_{def}$ <[s-rootl], ι>

121. BSR_0,{a,b,d,e},**V ∩ Pt**(<χ, ι>) =$_{def}$ <[stemg[7]], ι>

122. BSR_0,{c},**V ∩ Pt**(<χ, ι>) =$_{def}$ <[stemg[6]], ι>

123. BSR_0,{ɛ},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[steml[9]], ι>

For Brazilian Portuguese innovative past participles:

123+. BSR_0,{ɛ}, (**TruncPTCP**) ∩ **Pt**(<χ, ι>) =$_{def}$ <[stemk[9]], ι> *inserted*

124. BSR_0,{f},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[stemh[4]], ι>

For a variety with no active/passive distinction:

122. BSR_0,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[stemg[6]], ι>

123b. BSR_0,{c},**IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[steml[9]], ι> *edited*

123+. BSR_0,{c},**TruncPTCP** ∩ Pt(<χ, ι>) =$_{def}$ <[stemk[9]], ι> *inserted*

- Block 1 stem formation rules

125. SR_1,{a,b,f},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χ_i/r/], ι> for all i

126. SR_1,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/d/], ι>for all i

127. SR_1,{ɛ},**T-IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/t/], ι> for all i

128. SR_1,{ɛ},**S-IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/s/], ι> for all i

129. SR_1,{ɛ},**TruncPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi], ι>for all i

130. SR_1,{e},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χ_i/nd/], ι> for all i

131. SR_1,{d},**pplImp**(<χ, ι>) =$_{def}$ <γ, ι>, where Nar1,**V** ∩ **Pt**(<χ, c>) = <γ, c>

132. SR_1,{d},**gerImp**(<χ, ι>) =$_{def}$ <γ, ι>, where Nar1,**V** ∩ **Pt**(<χ, e>) = <γ, e>

For a variety with no active/passive distinction:

126. SR_1,{c},**V** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/d/], ι>for all i

127b. SR_1,{c},**T-IrregPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi/t/], ι> for all i *edited*

128b. SR_1,{c},**S-IrregPTCP** ∩**Pt**(<χ, ι>) =$_{def}$ <[χi/s/], ι> for all i *edited*

129b. SR_1,{c},**TruncPTCP** ∩ **Pt**(<χ, ι>) =$_{def}$ <[χi], ι>for all i *edited*

Stem selection rules

- Basic stem selection rules

133. SSR_0,←{}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemg [1]], σ>

134. where τ =← {tns: pres, md: subj} →,

SSR_0,τ,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemk [2]], σ>

↕

SSR_0,τ/{agr: {per: 1, num: sg},md: ind},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemk [2]], σ>

135. where τ =← {tns: past 2, md: ind} →,

SSR_0,τ,**PYTA** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemh [4]], σ>

↕

SSR_0,τ/{tns: plup,md: ind},**PYTA** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemh [4]], σ>

↕

SSR_0,τ/{tns: past 1,md: subj},**PYTA** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemh [4]], σ>

136. where τ =← {tns: pres, md: ind} →,

τ ≠← {agr: {per: 2, num: pl}} →

and ϱ = {md: subj}

SSR_0,τ,{IR}∩ **Pt**(<χ, σ>) =$_{def}$ <ψ, σ>, where Nar0,{IR}∩ **Pt**(<χ, σ/ϱ>) = <ψ, σ/ϱ>

137. SSR_0,←{tns: past 1,md: ind}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemn [1]], σ>

138. SSR_0,←{tns: past 1,md: ind}→,(**C2**∪**C3**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemn [3]], σ>

139. SSR_0,←{tns: past 1,md: ind}→,{SER} ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemn [10]], σ>

- Complex stem selection rules

140. SSR_0,←{vform: inf}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stema [10]], σ>

141. SSR_0,←{tns: fut,md: ind}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemb [10]], σ>

142. SSR_0,←{tns: cond}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemb [3]], σ>

143. SSR_0,{vform: ppl},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemc [10]], σ>

144. SSR_0,{vform: ppl}, (**IrregPTCP**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemɕ [8]], σ>

145. SSR_0,{vform: ppl, voice: active}, (**DoublePTCP**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemc [10]], σ>

For a variety with no active/passive distinction:

143b. SSR_0,{vform: ppl},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemc [8]], σ>

~~144b. SSR_0,{vform: ppl}, (**IrregPTCP**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemɕ [8]], σ>~~

~~145b. SSR_0,{vform: ppl, voice: active}, (**DoublePTCP**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemc [10]], σ>~~

146. SSR_0,{agr: {per: 2, num: pl},md: imp},**IrrImp** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemd [5]], σ>

147. SSR_0,{vform: ger},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[steme [12]], σ>

148. where α = 1,2,3, β = sg,pl,

SSR_0,{agr: {per: α, num: β},tns: fut,md: subj},**PYTA** ∩ **Pt**(<χ, σ>) =$_{def}$ <[stemf [10]], σ>

Whole word syncretism rules

149. where n = 0,1,2,

τ = ← {agr: {per: 1, num: sg}} but not of

{tns: pres, md: ind}, {tns: past 2, md: ind},

{tns: plup, md: ind}or {tns: fut, md: ind} →

and ϱ = {agr: {per: 3}}

RR_n,τ,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <ψ, σ>, where Nar_n,**V** ∩ **Pt**(<χ, σ/ϱ>) = <ψ, σ/ϱ>

150. where n = 0,1,2,

τ = {agr: {per: 2, num: sg}, md: imp}

and ϱ = {agr: {per: 3}, tns: pres, md: ind}

RR_n,τ,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <ψ, σ>, where Nar_n,**V** ∩ **Pt**(<χ, σ/ϱ>) = <ψ, σ/ϱ>

151. where n = 0,1,2,

τ = {agr: {per: 3, num: pl}, tns: past 2, md: ind}

and ϱ = {tns: plup}

$RR_{n},\tau,\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <\psi, \sigma>$, where $Nar_{n},\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma/\varrho>) = <\psi, \sigma/\varrho>$

152. where n = 0,1,2,

τ = {vform: ppl}

and ϱ = {vform: ger}

$RR_{n},\tau,\{vir\} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <\psi, \sigma>$, where $Nar_{n},\{vir\} \cap \mathbf{Pt}(<\chi, \sigma/\varrho>) = <\psi, \sigma/\varrho>$

153. where n = 0,1,2,

τ = {agr: {per: α, num: β}, tns: fut, md: subj}

and ϱ = {agr: {per: α, num: β}, vform: inf}

where α and β are any permitted values

$RR_{n},\tau,\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <\psi, \sigma>$, where $Nar_{n},\mathbf{V} \cap \mathbf{Pt}(<\chi, \varrho>) = <\psi, \varrho>$

Realisation rules

- Block 1 realisation rules

154. $RR_{1},\leftarrow\{tns: pres,md: ind\}\rightarrow,\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <[\chi_{i}/o/], \sigma>$ for all i

155. where τ = {agr: {per: 1, num: pl}, tns: pres, md: ind}

and ϱ = {agr: {per: 1, num: sg}}

$RR_{1},\tau,\{SER\} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <\psi, \sigma>$, where $Nar1,\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma/\varrho>) = <\psi, \sigma/\varrho>$

↕

$RR_{1},\tau/\{agr: \{per: 2\}\},\{SER\} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <\psi, \sigma>$, where

$Nar1,\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma/\varrho>) = <\psi, \sigma/\varrho>$

156. $RR_{1},\leftarrow\{tns: past\ 1,md: ind\}\rightarrow,\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <[\chi_{i}/va/], \sigma>$ for all i

157. where $\tau = \leftarrow$ {tns: past 1, md: ind} $\rightarrow$

and ϱ = {tns: cond},

$RR_{1},\tau,(\mathbf{C2}\cup\mathbf{C3}) \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <\psi, \sigma>$, where $Nar1,\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma/\varrho>) = <\psi, \sigma/\varrho>$

158. $RR_{1},\leftarrow\{tns: cond\}\rightarrow,\mathbf{V} \cap \mathbf{Pt}(<\chi, \sigma>) =_{def} <[\chi_{i}/a/], \sigma>$ for all i

159. RR_1,←{tns: past 1,md: subj}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/se/], σ> for all i

160. RR_1,{agr: {per: 1, num: sg},tns: past 2,md: ind},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/i/], σ> for all i

161. RR_1,←{agr: {per: 2},tns: past 2,md: ind}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/ste/], σ> for all i

162. RR_1,{agr: {per: 3, num: sg},tns: past 2,md: ind},(**C1****PYTA**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/o/], σ> for all i

163. RR_1,{agr: {per: 3, num: sg},tns: past 2,md: ind},(**V****PYTA**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/N/], σ> for all i

164. RR_1,←{tns: plup}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/ra/], σ> for all i

165. RR_1,←{agr: {per: 1},tns: fut,md: ind}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/é/], σ> for all i

166. where τ = {agr: {per: 2, num: pl}, tns: fut, md: ind}

and ϱ = {agr: {per: 1}}

RR_1,τ,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <ψ, σ>, where Nar1,**V** ∩ **Pt**(<χ, σ/ϱ>) = <ψ, σ/ϱ>

167. RR_1,←{agr: {per: 3},tns: fut,md: ind}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/á/], σ> for all i

168. where τ = {agr: {per: 2, num: sg}, tns: fut, md: ind}

and ϱ = {agr: {per: 3}}

RR_1,τ,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <ψ, σ>, where Nar1,**V** ∩ **Pt**(<χ, σ/ϱ>) = <ψ, σ/ϱ>

169. where α = 1,2,3,β = sg,pl,

τ = {agr: {per: α, num: β}, tns: pres, md: ind}

and ϱ = {tns: fut}

RR_1,τ,{HAVER} ∩ **Pt**(<χ, σ>) =$_{def}$ <ψ, σ>, where Nar1,**V** ∩ **Pt**(<χ, σ/ϱ>) = <ψ, σ/ϱ>

170. RR_1,{vform: ppl, agr: {gend: fem},**V** ∩ **Pt**(<χo, σ>) =$_{def}$ <[χ_i/a/], σ> for all i

- Block 2 realisation rules

171. RR_2,{agr: {per: 1, num: sg},tns: pres,md: ind},**1sgPI** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/u/], σ> for all i

172. RR_2,←{agr: {per: 2, num: sg}}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/s/], σ> for all i

173. RR_2,←{agr: {per: 2, num: pl},md: imp}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/i/], σ> for all i

174. RR_2,←{agr: {per: 1, num: pl}}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/mos/], σ> for all i

175. RR_2,←{agr: {per: 2, num: pl}}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/is/], σ> for all i

176. RR_2,{agr: {per: 2, num: pl},vform: inf},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/des/], σ> for all i

177. RR_2,←{agr: {per: 3, num: pl}}→,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/u/], σ> for all i

178. where n = 2,

τ is {agr: {per: 3, num: pl}, tns: pres, md: ind},

ϱ {agr: {per: 3, num: pl}, vform: inf}

RR_n,τ,(**C2**∪**C3**) ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/i/], σ> for all i

↕

RR_n,τ/{tns: past 1,md: subj},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/i/], σ> for all i

↕

RR_n,ϱ,**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/i/], σ> for all i

179. RR_2,{vform: ppl, agr: {num: pl},**V** ∩ **Pt**(<χ, σ>) =$_{def}$ <[χ_i/s/], σ> for all i

References

Aalberse, Suzanne, and Wessel Stoop. 2015. 'The Exceptional Loss of the Pronoun T'. *Journal of Pragmatics* 88: 190–201. https://doi.org/10.1016/j.pragma.2015.07.003.

Aaron Smith, K. 2012. 'New Perspectives, Theories and Methods: Frequency and Language Change'. In *English Historical Linguistics: An International Handbook*, by Alexander Bergs and Laurel Brinton, 2:1531–46. De Gruyter Mouton.

Aarts, Bas. 2004. 'Conceptions of Gradience in the History of Linguistics'. *Language Sciences* 26 (4): 343–89.

———. 2007. *Syntactic Gradience: The Nature of Grammatical Indeterminacy*. Oxford: Oxford University Press.

Ackerman, Farrell, and Gregory T. Stump. 2004. 'Paradigms and Periphrasis: A Study in Realization-Based Lexicalism'. In *Projecting Morphology*, edited by Louisa Sadler and Andrew Spencer, 111–57. Stanford: CSLI Publications.

Alencar, Leonel de. 2013. 'BrGram: Uma Gramática Computacional de Um Fragmento Do Português Brasileiro No Formalismo Da LFG'. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology. Fortaleza, Ceará, Brazil, October 20*, 24:183–88.

Ali, M. Said. (1908) 1957. *Dificuldades da língua portuguêsa: estudos e observacões*. 5. ed. Biblioteca brasileira de filologia, No. 1. Rio de Janeiro: Livraria Acadêmica.

———. (1908) 2008. *Dificuldades da língua portuguêsa: estudos e observacões*. 7. ed. Biblioteca brasileira de filologia, No. 1. Rio de Janeiro: Academia Brasileira de Letras: Biblioteca Nacional.

Alsina, Alex. Forthcoming. 'LFG and Romance Languages'. In *The Handbook of Lexical Functional Grammar*, by Mary Dalrymple. Empirically Oriented Theoretical Morphology and Syntax. Language Science Press.

Alsina, Alex, and Boban Arsenijević. 2012a. 'The Two Faces of Agreement'. *Language* 88 (2): 369–79. https://doi.org/10.1353/lan.2012.0039.

———. 2012b. 'There Is No Third Face of Agreement'. *Language* 88 (2): 388–89. https://doi.org/10.1353/lan.2012.0047.

Alvarez Blanco, María Rosario, Henrique Monteagudo Romero, and Xosé Luís Regueira. 1986. *Gramática Galega*. Biblioteca Básica da Cultura Galega. Vigo: Galaxia.

Amaral, Patrícia. 2019. 'Descriptive Pronouns'. In *Secondary Content*, 58–86. Brill.

Anderson, James M. 1973. *Structural Aspects of Language Change*. London: Longman.

Aronoff, Mark. 1994. *Morphology by Itself: Stems and Inflectional Classes*. Linguistic Inquiry Monographs 22. Cambridge, Mass.; London: MIT Press.

Asudeh, Ash, Mary Dalrymple, and Ida Toivonen. 2008. 'Constructions with Lexical Integrity: Templates as the Lexicon-Syntax Interface'. In *Proceedings of the LFG08 Conference*, edited by Miriam Butt and Tracy Holloway King, 68–88. Stanford: CSLI Publications.

Asudeh, Ash, and Gianluca Giorgolo. 2012. 'Flexible Composition for Optional and Derived Arguments'. In *Proceedings of the LFG12 Conference*, edited by Miriam Butt and Tracy Holloway King, 64–84. Stanford: CSLI Publications.

Baerman, Matthew. 2007a. 'Morphological Typology of Deponency'. In *Deponency and Morphological Mismatches*, edited by Matthew Baerman, Greville G. Corbett, Dunstan Brown, and Andrew Hippisley, 1–19. British Academy.

———. 2007b. 'Syncretism'. *Language and Linguistics Compass* 1 (5): 539–51. https://doi.org/10.1111/j.1749-818X.2007.00024.x.

Baerman, Matthew, Dunstan Brown, and Greville G. Corbett. 2005. *The Syntax-Morphology Interface: A Study of Syncretism*. Cambridge: Cambridge University Press.

Baerman, Matthew, Greville G. Corbett, Dunstan Brown, and Andrew Hippisley, eds. 2007. *Deponency and Morphological Mismatches*. British Academy.

Baker, Mark. 2003. *Lexical Categories: Verbs, Nouns and Adjectives*. Cambridge: Cambridge University Press.

Baker, Mark, and William Croft. 2017. 'Lexical Categories: Legacy, Lacuna, and Opportunity for Functionalists and Formalists'. *Annual Review of Linguistics* 3 (1): 179–97.

Baxter, Alan N. 1991. 'Portuguese as a Pluricentric Language'. In *Pluricentric Languages: Differing Norms in Different Nations*, edited by Michael Clyne, 11–44. Berlin: De Gruyter Mouton.

Baxter, Alan N. 1997. 'Creole-like Features in the Verb System of an Afro-Brazilian Variety of Portuguese'. In *The Structure and Status of Pidgins and Creoles*, edited by Arthur K. Spears and Donald Winford, 265–88. Creole Language Library 19. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Beard, Robert. 1995. *Lexeme-Morpheme Base Morphology: A General Theory of Inflection and Word Formation*. SUNY Series in Linguistics. Albany: State University of New York.

Bechara, Evanildo. 2015. *Moderna Gramática Portuguesa*. 38th ed. Rio de Janeiro: Editora Nova Fronteira.

Berman, Judith. 1999. 'Does German Satisfy the Subject Condition?' In *Proceedings of the LFG99 Conference*, edited by Miriam Butt and Tracy Holloway King, 20. Stanford: CSLI Publications.

———. 2003. *Clausal Syntax of German*. Stanford: CSLI Publications.

Bick, Eckhard. 2001. 'Portuguese Syntax: Teaching Manual'. 2001. https://visl.sdu.dk/visl/pt/info/syntaxmanual_0.html.

Biswas, Alexander A. 2017. 'Inflectional Periphrasis in LFG'. MPhil Thesis, University of Oxford.

———. 2019. 'The Passé Composé and Passé Surcomposé Tenses in French: An LFG Implementation of the Collocational Approach to Inflectional Periphrasis'. Presented at the 27th South of England LFG Meeting, February 9.

Blevins, James P., and Juliette Blevins. 2009. *Analogy in Grammar: Form and Acquisition*. Oxford Linguistics. Oxford: University Press.

Bonami, Olivier, and Ana R. Luís. 2015. 'Sur La Morphologie Implicative Dans La Conjugaison Du Portugais: Une Étude Quantitative'. *Mémoires de La Société de Linguistique de Paris*, 111–51.

Bonami, Olivier, and Pollet Samvelian. 2015. 'The Diversity of Inflectional Periphrasis in Persian'. *Journal of Linguistics* 51 (2): 327–82.

Bonami, Olivier, and Gregory Stump. 2016. 'Paradigm Function Morphology'. In *The Cambridge Handbook of Morphology*, edited by Andrew Hippisley and Gregory Stump, 449–81. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781139814720.017.

Borik, Olga. 2014. 'The Argument Structure of Long and Short Form Adjectives and Participles in Russian'. *Lingua*, SI: Aspect and Argument Structure of Adjectives and Participles, 149 (September): 139–65. https://doi.org/10.1016/j.lingua.2014.05.019.

Börjars, Kersti, Pauline Harries, and Nigel Vincent. 2016. 'Growing Syntax: The Development of a DP in North Germanic'. *Language (Baltimore)* 92 (1): e1–37. https://doi.org/10.1353/lan.2016.0002.

Börjars, Kersti, Rachel Nordlinger, and Louisa Sadler. 2019. *Lexical-Functional Grammar: An Introduction*. Cambridge: Cambridge University Press.

Börjars, Kersti, and Nigel Vincent. 2011. 'Grammaticalization and Directionality'. In *The Oxford Handbook of Grammaticalization*, edited by Bernd Heine and Heiko Narrog. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199586783.013.0014.

———. 2017. 'Lexical-Functional Grammar'. In *The Cambridge Handbook of Historical Syntax*, edited by Adam Ledgeway and Ian Roberts, 642–63. Cambridge: Cambridge University Press.

Bresnan, Joan. 1982. *The Mental Representation of Grammatical Relations*. Vol. 1. Cambridge, Mass.: MIT Press.

Bresnan, Joan, Ash Asudeh, Ida Toivonen, and Stephen Wechsler. 2015. *Lexical Functional Syntax*. 2nd ed. Chichester: John Wiley & Sons.
Bresnan, Joan, and Jonni M. Kanerva. 1989. 'Locative Inversion in Chichewa: A Case Study of Factorization in Grammar'. *Linguistic Inquiry* 20 (1): 1–50.
Bresnan, Joan, and Sam A. Mchombo. 1987. 'Topic, Pronoun, and Agreement in Chichewa'. *Language*, 741–82.
Brosman, Paul. 1963. 'Germanic "Waidanjan" in Romance'. *Romance Notes* 4 (2): 169–77.
Brown, Roger, and Albert Gilman. 1960. *The Pronouns of Power and Solidarity*. Berlin: De Gruyter Mouton.
Butt, Miriam, Mary Dalrymple, and Anette Frank. 1997. 'An Architecture for Linking Theory in LFG'. In *Proceedings of the LFG97 Conference*. Stanford: CSLI Publications.
Butt, Miriam, Christian Fortmann, and Christian Rohrer. 1996. 'Syntactic Analyses for Parallel Grammars: Auxiliaries and Genitive NPs'. In *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*. https://www.aclweb.org/anthology/C96-1032.
Butt, Miriam, and Tracy Holloway King. 2001. *Time over Matter: Diachronic Perspectives on Morphosyntax*. Studies in Constraint-Based Lexicalism. Stanford: CSLI Publications.
Butt, Miriam, Tracy Holloway King, Maria-Eugenia Niño, and Frédérique Segond. 1999. *A Grammar Writer's Cookbook*. Stanford: CSLI Publications.
Butt, Miriam, and Aditi Lahiri. 2013. 'Diachronic Pertinacity of Light Verbs'. *Lingua*, SI: Complex Predicates, 135 (October): 7–29. https://doi.org/10.1016/j.lingua.2012.11.006.
Butt, Miriam, María-Eugenia Niño, and Frédérique Segond. 1996. 'Multilingual Processing of Auxiliaries within LFG.' In *KONVENS*, 96:111–22.
Bybee, Joan L. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam; Philadelphia: John Benjamins Publishing Company.
Castro, Ivo. 2006. *Introdução à história do português*. 2a ed. Lisboa: Edições Colibri.
Chagas de Souza, Paulo. 2007. 'Athematic Participles in Brazilian Portuguese: A Syncretism in the Making'. *Acta Linguistica Hungarica* 54 (2): 119–28. https://doi.org/10.1556/aling.54.2007.2.1.
———. 2011. 'Particípios atemáticos no PB: Um processo paradigmático'. *ReVEL* 5. http://www.revel.inf.br/files/artigos/revel_esp_5_participios_atematicos_no_pb.pdf.
———. 2015. 'Particípios atemáticos do PB: paradigmas, sincretismo e diacronia'. In *Os Estudos Lexicais em Diferentes Perspectivas*, edited by Ieda Maria Alves and Ana Maria Ribeiro de Jesus, 29–57.
Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, Mass.: MIT Press.
Collins, Chris, and Paul Martin Postal. 2012. *Imposters: A Study of Pronominal Agreement*. Cambridge, Mass.: MIT Press.
Comrie, Bernard. 1975. 'Polite Plurals and Predicate Agreement'. *Language* 51 (2): 406–18. https://doi.org/10.2307/412863.
Cook, Manuela. 1997. 'Uma Teoria de Interpretação Das Formas de Tratamento Na Língua Portuguesa'. *Hispania* 80 (3): 451–64. https://doi.org/10.2307/345821.
———. 2013. 'Portuguese Pronouns and Other Forms of Address, from the Past into the Future—Structural, Semantic and Pragmatic Reflections'. *Journal of Lusophone Studies* 11.
Coppock, Elizabeth, and Stephen Wechsler. 2010. 'Less-Travelled Paths from Pronoun to Agreement: The Case of the Uralic Objective Conjugations'. In *Proceedings of the LFG10 Conference*, edited by Miriam Butt and Tracy Holloway King, 186. Stanford: CSLI Publications.
Corbett, Greville G. 1979. 'The Agreement Hierarchy'. *Journal of Linguistics* 15 (2): 203–24.
———. 1991. *Gender*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

———. 2006. *Agreement*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
———. 2009. 'Agreement'. In *Die Slavischen Sprachen/The Slavic Languages*. Berlin: De Gruyter.
Costa, Joao, and Sandra Pereira. 2013. 'A Gente: Pronominal Status and Agreement Revisited'. *The Linguistic Review* 30 (2): 161–84. https://doi.org/10.1515/tlr-2013-0006.
Crysmann, Berthold. 1997. 'Cliticization in European Portuguese Using Parallel Morpho-Syntactic Constraints'. In *Proceedings of the LFG97 Conference*, edited by Miriam Butt and Tracy Holloway King. Stanford: CSLI Publications.
Cunha, Celso Ferreira da. 1972. *Gramática do português contemporâneo*. 3. ed. Belo Horizonte: Editôra B. Alvares.
Cunha, Celso Ferreira da, and Luís F. Lindley Cintra. 1990. *Nova gramática do português contemporâneo*. 7th ed. Lisbon: Edições João Sá de Costa.
Cyrino, Sonia Maria Lazzarini. 1993. 'Observações Sobre a Mudança Diacrônica No Português Do Brasil: Objeto Nulo e Clíticos'. In *Português Brasileiro: Uma Viagem Diacrônica*, edited by Ian Roberts and Mary Aizawa Kato, 163–85. Campinas: Editora da UNICAMP.
Cyrino, Sonia Maria Lazzarini, and Gabriela Matos. 2016. 'Null Objects and VP Ellipsis in European and Brazilian Portuguese'. In *The Handbook of Portuguese Linguistics*, edited by W. Leo Wetzels, João Costa, and Sergio Menuzzi, 1st ed., 294–317. Wiley. https://doi.org/10.1002/9781118791844.ch29.
Da Tos, Martina. 2012. 'Il Participio Perfetto a Venezia: forme concorrenti'. *Quaderni di Lavoro ASIt* 14: 141–54.
Dalrymple, Mary. 2011. *Objects and Information Structure*. Cambridge Studies in Linguistics 131. Cambridge: Cambridge University Press.
———. 2015. 'Morphology in the LFG Architecture'. In *Proceedings of the LFG15 Conference*, edited by Miriam Butt and Tracy Holloway King. Stanford: CSLI Publications.
Dalrymple, Mary, Dag Haug, and John Lowe. 2018. 'Integrating LFG's Binding Theory with PCDRT'. *Journal of Language Modelling* 6 (1). https://doi.org/10.15398/jlm.v6i1.204.
Dalrymple, Mary, Ronald M. Kaplan, and Tracy Holloway King. 2004. 'Linguistic Generalizations over Descriptions'. In *Proceedings of the LFG04 Conference*, edited by Miriam Butt and Tracy Holloway King, 10. Stanford: CSLI Publications.
Dalrymple, Mary, John Lamping, and Vijay Saraswat. 1993. 'LFG Semantics via Constraints'. In *Sixth Conference of the European Chapter of the Association for Computational Linguistics*. Utrecht, The Netherlands: Association for Computational Linguistics. https://www.aclweb.org/anthology/E93-1013.
Dalrymple, Mary, John J. Lowe, and Louise Mycock. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press.
Dalrymple, Mary, and Louise Mycock. 2011. 'The Prosody-Semantics Interface'. In *Proceedings of the LFG11 Conference*, edited by Miriam Butt and Tracy Holloway King, 173–93. Stanford: CSLI Publications.
Davies, Mark. 2008. *A Frequency Dictionary of Portuguese: Core Vocabulary for Learners*. Routledge Frequency Dictionaries. New York: Routledge.
———. 2016. 'Corpus Do Português: One Billion Words, 4 Countries'. *Available Online at Http://Www.Corpusdoportugues.Org/Web-Dial/. (Web / Dialects)*.
Davies, Mark, and Michael Ferreira. 2006. 'Corpus Do Português: 45 Million Words, 1300s-1900s'. *Available Online at Http://Www.Corpusdoportugues.Org/Hist-Gen/. (Genre/Historical)*.
Déchaine, Rose-Marie, and Martina Wiltschko. 2002. 'Decomposing Pronouns'. *Linguistic Inquiry* 33 (3): 409–42.
———. 2003. 'On Pro-Nouns and Other "Pronouns"'. *From NP to DP* 1: 71–89.
Diessel, Holger. 2007. 'Frequency Effects in Language Acquisition, Language Use, and Diachronic Change'. *New Ideas in Psychology* 25 (2): 108–27.

———. 2012. '101. New Perspectives, Theories and Methods: Diachronic Change and Language Acquisition'. In *Handbücher Zur Sprach- Und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science 34.2*, edited by Alexander Bergs and Laurel Brinton. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110251609.1599.

Dikken, Marcel Den. 2001. '"Pluringulars", Pronouns and Quirky Agreement'. *The Linguistic Review* 18 (1). https://doi.org/10.1515/tlir.18.1.19.

Dixon, R. M. W. 2004. 'Adjective Classes in Typological Perspective'. In *Adjective Classes: A Cross-Linguistic Typology*, by R. M. W. Dixon and Alexandra Y. Aikhenvald. Oxford: Oxford University Press.

Duarte, Inês, and Gabriela Matos. 2000. 'Romance Clitics and the Minimalist Program'. In *Portuguese Syntax: New Comparative Studies*, edited by João Costa, 116–42. Oxford: Oxford University Press.

Duarte, Maria Eugênia Lammoglia. 1993. 'Do Pronome Nulo Ao Pronome Pleno: A Trajetória Do Sujeito No Português Do Brasil'. In *Português Brasileiro: Uma Viagem Diacrônica*, edited by Ian Roberts and Mary Aizawa Kato, 107–25. Campinas: Editora da UNICAMP.

———. 2017. 'Analyzing a Parametric Change in Brazilian Portuguese'. In *Studies on Variation in Portuguese*, edited by Pilar P. Barbosa, Maria da Conceição de Paiva, and Celeste Rodrigues, 234–53. Amsterdam: John Benjamins. https://benjamins.com/catalog/ihll.14.10dua.

Duarte, Maria Eugênia Lammoglia, and Humberto Soares da Silva. 2016. 'Microparametric Variation in Spanish and Portuguese: The Null Subject Parameter and the Role of the Verb Inflectional Paradigm'. In *The Morphosyntax of Portuguese and Spanish in Latin America*, by Mary Aizawa Kato and Francisco Ordóñez, 1–26. Oxford: Oxford University Press.

Duarte, Vânia Maria do Nascimento. n.d. 'Você: Segunda Ou Terceira Pessoa Do Discurso?' Brasil Escola. Accessed 29 April 2021. https://brasilescola.uol.com.br/gramatica/voce-segunda-ou-terceira-pessoa-discurso.htm.

Dubert, Francisco, and Charlotte Galves. 2016. 'Galician and Portuguese'. In *The Oxford Guide to the Romance Languages*, by Adam Ledgeway and Martin Maiden, 411–46. Oxford: Oxford University Press.

Embick, David. 2004. 'On the Structure of Resultative Participles in English'. *Linguistic Inquiry* 35 (3): 355–92. https://doi.org/10.1162/0024389041402634.

Embick, David, and Morris Halle. 2005. 'On the Status of Stems in Morphological Theory'. In *Romance Languages and Linguistic Theory 2003: Selected Papers from 'Going Romance' 2003*, edited by Twan Geerts, Ivo vanGinneken, and Haike Jacobs, 270:37–62. Current Issues in Linguistic Theory. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/cilt.270.03emb.

Esher, Louise. 2017. 'Morphome Death and Transfiguration in the History of French'. *Journal of Linguistics* 53 (1): 51–84. https://doi.org/10.1017/S0022226715000468.

———. 2020. 'Syncretism and Metamorphomes in the Diachrony of Lemosin Varieties'. In *Variation and Change in Gallo-Romance Grammar*, edited by Sam Wolfe and Martin Maiden. Oxford: Oxford University Press. https://hal.archives-ouvertes.fr/hal-02929770.

Fertig, David. 2013. *Analogy and Morphological Change*. Edinburgh Historical Linguistics. Edinburgh: Edinburgh University Press.

Findlay, Jamie Yates. 2016. 'Mapping Theory without Argument Structure'. *Journal of Language Modelling* 4 (2): 293-338-293–338. https://doi.org/10.15398/jlm.v4i2.171.

Finkel, Raphael. 2010. 'What Your Teacher Told You Is True: Latin Verbs Have Four Principle Parts'. In *Changing the Center of Gravity*, edited by Melissa Terras and Gregory Crane, 263–96. Gorgias Press. https://doi.org/10.31826/9781463219222-014.

———. 2016. 'Computer-Based Tools for Word and Paradigm Computational Morphology'. In *Oxford Research Encyclopedia of Linguistics*, by Raphael Finkel. Oxford: Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.162.

Fong, Suzana. 2019. 'Regular and Athematic Participles in Brazilian Portuguese'. In *MorPhun*, 20. http://sznfong.scripts.mit.edu/site/wp-content/uploads/2019/09/participle-BP-LSRL-49-paper.pdf.

Forst, Martin. 2011. 'Computational Aspects of Lexical Functional Grammar'. *Language and Linguistics Compass* 5 (1): 1–18. https://doi.org/10.1111/j.1749-818X.2010.00261.x.

Francis, Elaine J., and Laura A. Michaelis, eds. 2003. *Mismatch: Form-Function Incongruity and the Architecture of Grammar*. Stanford: CSLI Publications.

Frank, Anette, and Annie Zaenen. 2002. 'Tense in LFG: Syntax and Morphology'. *How We Say When It Happens: Contributions to the Theory of Temporal Reference in Natural Language*, 17–52.

Galves, Charlotte. 1993. 'O Enfraquecimento Da Concordância No Português Brasileiro'. In *Português Brasileiro: Uma Viagem Diacrônica*, edited by Ian Roberts and Mary Aizawa Kato, 387–408. Campinas: Editora da UNICAMP.

Galves, Charlotte, and Anthony Kroch. 2016. 'Main Syntactic Changes from a Principle-and-Parameters View'. In *The Handbook of Portuguese Linguistics*, edited by W. Leo Wetzels, João Costa, and Sergio Menuzzi, 1st ed., 487–503. Wiley. https://doi.org/10.1002/9781118791844.ch27.

Galves, Charlotte, M. Aparecida T. Moraes, and Ilza Ribeiro. 2005. 'Syntax and Morphology in the Placement of Clitics in European and Brazilian Portuguese'. *Journal of Portuguese Linguistics* 4 (2).

Gibert Sotelo, Elisabeth. 2017. 'Source and Negative Prefixes: On the Syntax-Lexicon Interface and the Encoding of Spatial Relations'. PhD Thesis, Girona: Universitat de Girona.

Girard, Jean-Yves. 1987. 'Linear Logic'. *Theoretical Computer Science* 50 (1): 1–101. https://doi.org/10.1016/0304-3975(87)90045-4.

Grandi, Nicola. 2015. *Edinburgh Handbook of Evaluative Morphology*. Edinburgh University Press.

Groot, Casper de. 2009. 'Identifiability and Verbal Cross-Referencing Markers in Hungarian'. *Linguistics* 47 (4): 997–1019. https://doi.org/10.1515/LING.2009.035.

Halle, Morris, and Alec Marantz. 1993. *Distributed Morphology and the Pieces of Inflection. Hale, K. & SJ Keyser (Eds.), The View from Building 20*. Cambridge, Mass.: MIT Press.

———. 1994. 'Some Key Features of Distributed Morphology'. *MIT Working Papers in Linguistics* 21 (275): 88.

Harrison, Rebecca. n.d. 'Latin 4th Principal Part Stem Types'. Cogitatorium. Accessed 27 November 2020. http://rharriso.sites.truman.edu/latin-4th-principal-part-stem-types/#-sus.

Haug, Dag. 2014. 'Partial Dynamic Semantics for Anaphora: Compositionality without Syntactic Coindexation'. *Journal of Semantics* 31 (4): 457–511. https://doi.org/10.1093/jos/fft008.

Haug, Dag, and Tatiana Nikitina. 2016. 'Feature Sharing in Agreement'. *Natural Language & Linguistic Theory* 34 (3): 865–910. https://doi.org/10.1007/s11049-015-9321-9.

Hautli, Annette, Özlem Çetinoğlu, and Josef van Genabith. 2010. 'Closing the Gap between Stochastic and Rule-Based Grammars'. In *Proceedings of the LFG10 Conference*, edited by Miriam Butt and Tracy Holloway King, 270–89. Stanford: CSLI Publications. http://web.stanford.edu/group/cslipublications/cslipublications/LFG/15/papers/lfg10hautlietal.pdf.

Hautli, Annette, and Tracy Holloway King. 2009. 'Adapting Stochastic LFG Input for Semantics'. In *Proceedings of the LFG09 Conference*, edited by Miriam Butt and Tracy Holloway King, 357–77. Stanford: CSLI Publications. http://web.stanford.edu/group/cslipublications/cslipublications/LFG/14/papers/lfg09hautliking.pdf.

Heine, Bernd, and Kyung-An Song. 2011. 'On the Grammaticalization of Personal Pronouns'. *Journal of Linguistics* 47 (3): 587–630.

Holm, John. 1994. 'A Semi-Crioulização Do Português Vernáculo Do Brasil: Evidência de Contacto Nas Expressões Idiomáticas'. *Papia: Revista de Crioulos de Base Ibérica* 3 (2).

Hristov, Bozhil. 2013. 'Defacing Agreement'. In *Proceedings of the LFG13 Conference*, edited by Miriam Butt and Tracy Holloway King, 335–55. CSLI Publications.

Ilari, Rodolfo. 2013. 'O Português No Contexto Das Línguas Românicas'. In *Gramática Do Português*, by Eduardo Buzaglo Paiva Raposo, Maria Fernanda Bacelar do Nascimento, Maria Antónia Coelho da Mota, Luísa Segura, and Amália Mendes, 1:49–66. Lisbon: Fundação Calouste Gulbenkian.

Inverno, Liliana. 2011. 'Contact-Induced Restructuring of Portuguese Morphosyntax in Interior Angola'. Doctoral thesis, Universidade de Coimbra.

Ishiyama, Osamu. 2019. *Diachrony of Personal Pronouns in Japanese: A Functional and Cross-Linguistic Perspective*. John Benjamins Publishing Company.

Jensen, John B. 1981. 'Forms of Address in Brazilian Portuguese: Standard European or Oriental Honorifics?' In *From Linguistics to Literature: Romance Studies Offered to Francis M. Rogers*, edited by Bernard H. Bichakjian, 45–66. Amsterdam: John Benjamins Publishing Company.

Johnson, Mark. 2000. 'Stochastic Lexical-Functional Grammars'. Presented at the LFG00 Conference, July. http://web.science.mq.edu.au/~mjohnson/papers/lfg00-slides.pdf.

Kamp, Hans. (1981) 2002. 'A Theory of Truth and Semantic Representation'. In *Formal Semantics: The Essential Readings*, edited by Paul Portner and Barbara H. Partee, 189–222. Oxford: Blackwell.

Kaplan, Ronald M. 1987. 'Three Seductions of Computational Psycholinguistics'. In *Linguistic Theory and Computer Applications*, edited by P. Whitelock, H. Somers, P. Bennett, R. Johnson, and M. McGee Wood, 149–88. London: Academic Press.

———. 2018. 'Formal Aspects of Underspecified Features'. In *Tokens of Meaning: Papers in Honor of Lauri Karttunen*, edited by Cleo Condoravdi and Tracy Holloway King, 349–69. CSLI Publications.

Kaplan, Ronald M., and Joan Bresnan. 1982. 'Lexical-Functional Grammar: A Formal System for Grammatical Representation'. In *The Mental Representation of Grammatical Relations*, by Joan Bresnan, 173–281. Cambridge, Mass.: MIT Press.

Kibort, Anna. 2007. 'Extending the Applicability of Lexical Mapping Theory'. In *Proceedings of the LFG07 Conference*, edited by Miriam Butt and Tracy Holloway King, 250–70.

———. 2009. 'Passive and Passive-like Constructions in English and Polish'. PhD Thesis, Cambridge: University of Cambridge.

Laca, Brenda. 2004. 'Romance "Aspectual" Periphrases: Eventuality Modification versus "Syntactic" Aspect'. In *The Syntax of Time*, edited by Jacqueline Guéron and Jacqueline Lecarme, 425–40. Cambridge, Mass.: MIT Press.

Lara Bermejo, Víctor. 2015. 'Los Tratamientos de 2pl En Andalucia Occidental y Portugal: Estudio Geo- y Sociolingüístico de Un Proceso de Gramaticalización'. Doctoral thesis, Universidade Autónoma de Madrid.

Lara Bermejo, Víctor, and Ana Guilherme. 2018. 'The Politeness of Você in European Portuguese'. *Studies in Hispanic and Lusophone Linguistics* 11 (2): 337–66. https://doi.org/10.1515/shll-2018-0012.

Laurent, Richard. 1999. *Past Participles from Latin to Romance*. University of California Publications in Linguistics 133. Berkeley; London: University of California Press.

Ledgeway, Adam. 2016. 'The Dialects of Southern Italy'. In *The Oxford Guide to the Romance Languages*, edited by Adam Ledgeway and Martin Maiden. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199677108.003.0016.

Levin, Lorraine S. 1986. 'Operations on Lexical Forms: Unaccusative Rules in Germanic Languages'. PhD Thesis, Massachusetts Institute of Technology.

Liao, Wei-Tai. 2010. 'An LFG Account of Empty Pronouns in Mandarin Chinese'. DPhil Thesis, University of Oxford.

Lightfoot, David. 1991. *How to Set Parameters: Arguments from Language Change*. Cambridge, Mass.: MIT Press.

Lobato, Lucia. 1999. 'Sobre a forma do particípio do português e o estatuto dos traços formais'. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada* 15 (1). https://doi.org/10.1590/S0102-44501999000100005.

Lopes, Célia Regina dos Santos. 2003. *A inserção de a gente no quadro pronominal do português*. Lingüística iberoamericana 18. Madrid; Frankfurt am Main: Iberoamericana; Vervuert.

Lopes, Célia Regina dos Santos, and Maria Teresa Brocardo. 2016. 'Main Morphosyntactic Changes and Grammaticalization Processes'. In *The Handbook of Portuguese Linguistics*, by W. Leo Wetzels, Sergio Menuzzi, and João Costa, 598–1h. Wiley Online Library.

Lopes, João Antunes. 1987. *Dicionário de Verbos*. Porto, Lello & Irmão.

Lowe, John J. 2016. 'Participles, Gerunds and Syntactic Categories'. In *Proceedings of the Joint 2016 Conference on Head-Driven Phrase Structure Grammar and Lexical Functional Grammar*, 401–21. Stanford: CSLI Publications.

———. 2019. 'Mixed Projections and Syntactic Categories'. *Journal of Linguistics* 56 (2): 1–43. https://doi.org/10.1017/S0022226719000100.

Lucchesi, Dante. 2001. 'As duas grandes vertentes da história sociolingüística do Brasil (1500-2000)'. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada* 17 (1): 97–130. https://doi.org/10.1590/S0102-44502001000100005.

———. 2017. 'A periodização da história sociolinguística do Brasil'. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada* 33 (2): 347–82. https://doi.org/10.1590/0102-445067529349614964.

Luís, Ana R. 2004. 'Clitics as Morphology'. PhD Thesis, University of Essex.

Luís, Ana R., and Ryo Otoguro. 2004. 'Proclitic Contexts in European Portuguese and Their Effect on Clitic Placement'. In *Proceedings of the LFG04 Conference*, edited by Miriam Butt and Tracy Holloway King, 334–52. Stanford: CSLI Publications.

———. 2005. 'Morphological and Syntactic Well-Formedness: The Case of European Portuguese Proclitics'. In *Proceedings of the LFG05 Conference*, edited by Miriam Butt and Tracy Holloway King, 253–70. Stanford: CSLI Publications.

Luís, Ana R., and Louisa Sadler. 2003. 'Object Clitics and Marked Morphology'. In *Empirical Issues in Formal Syntax and Semantics 4*, by Claire Beyssade, Olivier Bonami, Patricia Cabredo Hofherr, and Francis Corblin, 133–53. Paris: Presses Universitaires de Paris-Sorbonne.

Madeira, Ana Maria. 1992. 'On Clitic Placement in European Portuguese'. *UCL Working Papers in Linguistics* 4: 95–122.

Maiden, Martin. 2005. 'Morphological Autonomy and Diachrony'. In *Yearbook of Morphology 2004*, edited by Geert Booij and Jaap van Marle, 137–75. Yearbook of Morphology. Dordrecht: Springer. https://doi.org/10.1007/1-4020-2900-4_6.

———. 2011. 'Morphological Persistence'. In *The Cambridge History of the Romance Languages*, by John Charles Smith, Martin Maiden, and Adam Ledgeway, 1 (Structures):155–215. Cambridge: Cambridge University Press.

———. 2013. 'The Latin 'Third Stem'and Its Romance Descendants'. *Diachronica* 30 (4): 492–530.

———. 2016a. 'Inflectional Morphology'. In *The Oxford Guide to the Romance Languages*, edited by Adam Ledgeway and Martin Maiden, 497–512. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199677108.003.0027.

———. 2016b. 'Morphomes'. In *The Oxford Guide to the Romance Languages*, edited by Adam Ledgeway and Martin Maiden, 708–22. Oxford: Oxford University Press.

———. 2018. *The Romance Verb: Morphomic Structure and Diachrony*. Oxford: Oxford University Press.

Marcotulio, Leonardo, Juliana Vianna, and Célia Regina dos Santos Lopes. 2013. 'Agreement Patterns with a Gente in Portuguese'. *Journal of Portuguese Linguistics* 12 (2): 125–49. https://doi.org/10.5334/jpl.70.

Martins, Ana Maria. 2002. 'Mudança sintáctica e história da língua portuguesa'. In *História da Língua e História da Gramática: Actas do Encontro*, edited by Brian Franklin Head, 251–97. Braga: Universidade do Minho.

———. 2016. 'Introdução: O português numa perspetiva diacrónica e comparativa'. In *Manual de linguística portuguesa*, edited by Ana Maria Martins and Ernestina Carrilho, 1–40. Berlin: De Gruyter. https://doi.org/10.1515/9783110368840-003.

Matthews, P. H. 2007. *The Concise Oxford Dictionary of Linguistics*. 2nd ed. Oxford Paperback Reference. Oxford: Oxford University Press.

McCarren, Vincent P., and Mary Ann Ritter. 2020. 'An Edition of the Letters L, M, N and O of the Medulla Grammatice Stonyhurst MS. 15 (A.1.10)'. University of Michigan. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/143140/Medulla_Grammatice_LMNO.pdf?sequence=10.

McFadden, Thomas. 2007. 'Auxiliary Selection'. *Language and Linguistics Compass* 1 (6): 674–708. https://doi.org/10.1111/j.1749-818X.2007.00034.x.

Melchin, Paul B, Ash Asudeh, and Dan Siddiqi. 2020. 'Ojibwe Agreement in Lexical-Realizational Functional Grammar'. In *Proceedings of the LFG20 Conference*, edited by Miriam Butt and Ida Toivonen, 268–88. Stanford: CSLI Publications.

Miara, Fernanda Lima Jardim. 2013. 'Particípios Duplos: Usos, Desusos e Alguns "Intrusos"'. PhD Thesis, Florianópolis: Universidade Federal de Santa Catarina. https://repositorio.ufsc.br/bitstream/handle/123456789/107342/319589.pdf?seq.

Mittendorf, Ingo, and Louisa Sadler. 2005. 'Numerals, Nouns and Number in Welsh NPs'. In *Proceedings of the LFG05 Conference*, edited by Miriam Butt and Tracy Holloway King, 294–312. Stanford: CSLI Publications.

Móia, Telmo. 2004. 'Algumas Áreas Problemáticas Para a Normalização Linguística - Disparidades Entre o Uso e Os Instrumentos de Normalização'. In *Actas Do XX Encontro Nacional Da Associação Portuguesa de Linguística*, 109–25. Lisbon: Associação Portuguesa de Linguística.

Morgan, Terrell A., and Scott A. Schwenter. 2019. 'Asymmetry in Second Person Plural Reference: The Seemingly Intractable Cases of Romanian and European Portuguese'. In *INAR 05*. University of Sheffield. https://inarweb.files.wordpress.com/2019/06/ta.-morgan-s.-schwenter-2019.pdf.

Müller, Max. (1862) 2010. *Lectures on The Science of Language*. Gutenburg EBook. New York: Charles Scribner. http://www.gutenberg.org/files/32856/32856-pdf.pdf.

Mycock, Louise J. 2006. 'The Typology of Constituent Questions: A Lexical-Functional Grammar Analysis of 'Wh'-Questions.' PhD Thesis, University of Manchester.

Nevins, Andrew. 2015. 'Productivity and Portuguese Morphology: How Experiments Enable Hypothesis-Testing'. *Romance Languages and Linguistic Theory: Selected Papers from 'Going Romance' Amsterdam 2013*, no. 8: 175–201.

Nevins, Andrew, and Cilene Rodrigues. 2014. 'Athematic participle choice is predicted by argument structure'. *Revista da ABRALIN* 13 (2). https://doi.org/10.5380/rabl.v13i2.39737.

Nikolaeva, Irina, and Andrew Spencer. 2019. *Mixed Categories: The Morphosyntax of Noun Modification*. Cambridge Studies in Linguistics. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108233903.

Norde, Muriel. 2009. *Degrammaticalization*. Oxford: Oxford University Press.

———. 2010. 'Three Common Controversies'. In *Grammaticalization: Current Views and Issues*, edited by Katerina Stathi, Elke Gehweiler, and Ekkehard König, 119:123–50. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Nunes, Jairo. 1993. 'Direção de Cliticização, Objeto Nulo e Pronome Tônico Na Posição de Objeto Em Português Brasileiro'. In *Português Brasileiro: Uma Viagem Diacrônica*, edited by Ian Roberts and Mary Aizawa Kato, 207–22. Campinas: Editora da UNICAMP.

Nwoye, Onuigbo G. 1992. 'Linguistic Politeness and Socio-Cultural Variations of the Notion of Face'. *Journal of Pragmatics* 18 (4): 309–28. https://doi.org/10.1016/0378-2166(92)90092-P.

O'Connor, Robert. 2006. 'Information Structure in Lexical Functional Grammar: The Discourse-Prosody Correspondence.' PhD Thesis, Manchester: University of Manchester.

Oliveira, Fátima. 2013. 'Tempo Verbal'. In *Gramática Do Português*, by Eduardo Buzaglo Paiva Raposo, Maria Fernanda Bacelar do Nascimento, Maria Antónia Coelho da Mota, Luísa Segura, and Amália Mendes, 1:509–56. Lisbon: Fundação Calouste Gulbenkian.

O'Neill, Paul, and Gladis Massini-Cagliari. 2019. 'Linguistic Prejudice and Discrimination in Brazilian Portuguese and beyond: Suggestions and Recommendations'. *Journal of Language and Discrimination* 3 (1): 32–62.

Packard Humanities Institute. 2015. 'Classical Latin Texts: A Resource Prepared by the Packard Humanities Institute'. 2015. https://latin.packhum.org/index.

Pagotto, Emilio G. 1993. 'Clíticos, Mudança e Seleção Natural'. In *Português Brasileiro: Uma Viagem Diacrônica*, edited by Ian Roberts and Mary Aizawa Kato, 185–206. Campinas: Editora da UNICAMP.

Parkvall, Mikael. 2000. 'The Alleged Creole Past of Vernacular Brazilian Portuguese'. In *Crioulos de Base Portuguesa*, edited by Ernesto d'Andrade, Dulce Pereira, and Maria Antónia Coelho da Mota. Lisbon: Associação Portuguesa de Linguística.

Parsons, Terence. 1990. *Events in the Semantics of English*. Cambridge, Mass.: MIT Press.

Pollard, Carl Jesse, and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Studies in Contemporary Linguistics. Stanford, Chicago, London: University of Chicago Press and CSLI Publications.

Popova, Gergana, and Andrew Spencer. 2012. 'Relatedness in Periphrasis: A Paradigm-Based Perspective'. In *Periphrasis: The Role of Syntax and Morphology in Paradigms*, edited by Marina Chumakina and Greville G. Corbett. British Academy.

Porto Editora. 2003a. 'A Gente Vai Ou a Gente Vamos?' In *Dicionário Infopédia Da Língua Portuguesa*. Porto. https://www.infopedia.pt/$a-gente-vai-ou-a-gente-vamos.

———. 2003b. 'Ganhar'. In *Dicionário Infopédia Da Língua Portuguesa*. Porto. https://www.infopedia.pt/dicionarios/lingua-portuguesa/ganhar.

———. 2003c. 'Gastar'. In *Dicionário Infopédia Da Língua Portuguesa*. Porto. https://www.infopedia.pt/dicionarios/lingua-portuguesa/gastar.

Pountain, Christopher J. 2003. 'Pragmatic and Structural Reflections on the Expression of the Second Person Notion in Romance, with Special Reference to Spanish and Portuguese'. *Bulletin of Spanish Studies* 80 (2): 145–60. https://doi.org/10.1080/14753820302029.

Raposo, Eduardo Buzaglo Paiva. 2013. 'Pronomes'. In *Gramática Do Português*, by Eduardo Buzaglo Paiva Raposo, Maria Fernanda Bacelar do Nascimento, Maria Antónia Coelho da Mota, Luísa Segura, and Amália Mendes, 1:883–920. Lisbon: Fundação Calouste Gulbenkian.

Raposo, Eduardo Buzaglo Paiva, Maria Fernanda Bacelar do Nascimento, Maria Antónia Coelho da Mota, Luísa Segura, and Amália Mendes. 2013. *Gramática do Português*. Vol. 1. 3 vols. Lisbon: Fundação Calouste Gulbenkian.

Redenbarger, Wayne J. 1976. 'Vowel Lowering and I-Epenthesis in Classical Latin'. *Current Studies in Romance Linguistics, Georgetown University Press, Washington, DC* 16.

Robins, Robert Henry. 1959. 'In Defence of WP'. *Transactions of the Philological Society* 58 (1): 116–44.

Sadler, Louisa. 1997. 'Clitics and the Structure-Function Mapping'. In *Proceedings of the LFG97 Conference*, edited by Miriam Butt and Tracy Holloway King, 16. Stanford: CSLI Publications. http://web.stanford.edu/group/cslipublications/cslipublications/LFG/LFG2-1997/lfg97sadler.pdf.

Sadler, Louisa, and Rachel Nordlinger. 2018. 'Morphology in Lexical-Functional Grammar and Head-Driven Phrase Structure Grammar'. In *The Oxford Handbook of Morphological Theory*, edited by Jenny Audring and Francesca Masini. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199668984.013.17.

Sadler, Louisa, and Andrew Spencer. 2001. 'Syntax as an Exponent of Morphological Features'. In *Yearbook of Morphology 2000*, 71–96. Dordrecht: Springer.

———. 2004. *Projecting Morphology*. Stanford: CSLI Publications.

Scher, Ana Paula, Marcus Vinicius Lunguinho, and Aline Garcia Rodero-Takahira. 2013. 'Innovative Participles in Brazilian Portuguese'. *Revista Linguí∫tica* 9 (1). https://revistas.ufrj.br/index.php/rl/article/view/4568/0.

———. 2014. 'Voice (a)symmetries and innovative participles in Brazilian Portuguese'. *Cadernos de Estudos Lingüísticos* 56 (1): 45–64. https://doi.org/10.20396/cel.v56i1.8636524.

Scherre, Maria Marta Pereira, and Maria Eugênia Lammoglia Duarte. 2016. 'Main Current Processes of Morphosyntactic Variation'. In *The Handbook of Portuguese Linguistics*, edited by W. Leo Wetzels, João Costa, and Sergio Menuzzi, 1st ed., 526–44. Wiley. https://doi.org/10.1002/9781118791844.ch29.

Schwenter, Scott A., Mark Hoff, Eleni Christodulelis, Chelsea Pflum, and Ashlee Dauphinais. 2019. 'Variable Past Participles in Portuguese Perfect Constructions'. *Language Variation and Change* 31 (1): 69–89. https://doi.org/10.1017/S0954394519000036.

Schwenter, Scott A., and Gláucia V. Silva. 2003. 'Anaphoric Direct Objects in Spoken Brazilian Portuguese: Semantics and Pragmatics'. *Revista Internacional de Lingüística Iberoamericana* 1 (2): 99–123.

Segura, Luísa. 2013. 'Geografia Da Lingua Portuguesa'. In *Gramática Do Português*, by Eduardo Buzaglo Paiva Raposo, Maria Fernanda Bacelar do Nascimento, Maria Antónia Coelho da Mota, Luísa Segura, and Amália Mendes, 1:71–81. Lisbon: Fundação Calouste Gulbenkian.

Silva, Gláucia V. 2001. *Word Order in Brazilian Portuguese*. Berlin: Walter de Gruyter.

Smith, John Charles. 2010. 'Change and Continuity in Form-Function Relationships'. In *The Cambridge History of the Romance Languages*, edited by Martin Maiden, John Charles Smith, and Adam Ledgeway, 1st ed., 268–317. Cambridge University Press. https://doi.org/10.1017/CHOL9780521800723.008.

Spencer, Andrew. 2012. 'Identifying Stems'. *Word Structure* 5 (1): 88–108. https://doi.org/10.3366/word.2012.0021.

———. 2013. *Lexical Relatedness: A Paradigm-Based Model*. Oxford: Oxford University Press.

———. 2017. 'Participles: Inflectional Paradigms, Derivational Paradigms or Something Else?' In *ParadigMo*. http://w3.erss.univ-tlse2.fr/ParadigMo2017/diapos/Spencer_ParadigMo_2017_slides.pdf.

Spencer, Andrew, and Ana Luís. 2012. 'The Canonical Clitic'. In *Canonical Morphology and Syntax*, edited by Dunstan Brown, Marina Chumakina, and Greville G. Corbett. Oxford: Oxford University Press.

Spencer, Andrew, and Gergana Popova. 2015. *Periphrasis and Inflection*. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199591428.013.9.

Steele, Susan. 1995. 'Towards a Theory of Morphological Information'. *Language* 71 (2): 260–309. https://doi.org/10.2307/416164.

Stewart, Thomas, and Gregory Stump. 2007. 'Paradigm Function Morphology and the Morphology–Syntax Interface'. In *The Oxford Handbook of Linguistic Interfaces*, edited by Gillian Ramchand and Charles Reiss, 420. Oxford: Oxford University Press.

Stovicek, Thomas William. 2010. 'A Developmental History of the Hispano-Romance Verb Conjugations'. PhD Thesis, The Ohio State University.

Stump, Gregory T. 2001. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

———. 2015. 'Inflection Classes'. In *The Oxford Handbook of Inflection*. Oxford: Oxford University Press.

———. 2016. *Inflectional Paradigms: Content and Form at the Syntax-Morphology Interface*. Cambridge: Cambridge University Press.

Sugamoto, Nobuko. 1989a. 'Pronominality: A Noun-Pronoun Continuum'. In *Linguistic Categorization*, by Roberta Corrigan, Fred R. Eckman, and Michael Noonan, 267–91. Philadelphia, Pennsylvania: John Benjamins Publishing Company.

———. 1989b. 'The Pronominality of Referential Expressions'. PhD Thesis, United States -- California: University of California, Los Angeles. https://www.proquest.com/docview/303774011/abstract/76238091F364809PQ/1.

Svensen, Kristina. 2013. 'Los dobles participios en español: Estudio de corpus'. PhD Thesis, Universitetet i Tromsø. http://hdl.handle.net/10037/5279.

Sweet, Henry. 1888. *A History of English Sounds from the Earliest Period: With Full Word-Lists*. Vol. 11. Clarendon Press.

Szymanek, Bogdan. 1988. *Categories and Categorization in Morphology*. Katolicki Uniwersytet Lubelski. Wydział Nauk Humanistycznych. Rozprawa Habilitacyjna. Lublin: Redakcja Wydawnictw Katolickiego Uniwersytetu Lubelskiego.

Tarallo, Fernando. 1993. 'Sobre a Alegada Origem Crioula Do Português Brasileiro: Mudanças Sintáticas Aleatórias'. In *Português Brasileiro: Uma Viagem Diacrônica*, edited by Ian Roberts and Mary Aizawa Kato, 35–68. Campinas: Editora da UNICAMP.

Taylor, Catherine. 2012. 'Formal Relationships in the Paradigm: A Functional Approach with a Focus on the Romance Verb'. PhD Thesis, University of Essex.

Taylor, Michael. 2009. 'On the Pronominal Status of Brazilian Portuguese a Gente'. *NYU Working Papers in Linguistics* 2: 36.

Teixeira da Silva, Inaciane. 2008. 'O uso do particípio em formações verbais no português do sul do Brasil'. Master's Dissertation, Universidade Federal do Rio Grande do Sul. https://lume.ufrgs.br/handle/10183/17213.

Thomas, Amanda L R. 2018. 'Accounting for Double Past Participle Forms in Romance'. MPhil Thesis, University of Oxford.

Thomé-Williams, Ana Clotilde. 2004. 'Sociolinguistic Aspects of Forms of Address in Portugal and Brazil: Tu or Você'. *Intercultural Communication Studies* 13 (3): 85–99.

Thornton, Anna M. 2011. 'Overabundance (Multiple Forms Realizing the Same Cell): A Non-Canonical Phenomenon in Italian Verb Morphology'. In *Morphological Autonomy: Perspectives from Romance Inflectional Morphology*, edited by Martin Maiden, John Charles Smith, Maria Goldbach, and Marc-Olivier Hinzelin, 358–81. Oxford: Oxford University Press.

Toivonen, Ida. 2001. 'Language Change, Lexical Features and Finnish Possessors'. In *Time over Matter: Diachronic Perspectives on Morphosyntax*, edited by Miriam Butt and Tracy Holloway King, 209–25. Stanford: CSLI Publications.

Traugott, Elizabeth Closs. 2001. 'Legitimate Counterexamples to Unidirectionality'. Freiburg University, October 17.

Vieira, Sílvia Rodrigues, and Sílvia Figueiredo Brandão. 2017. 'Agreement in Portuguese'. In *Studies on Variation in Portuguese*, edited by Pilar P. Barbosa, Maria da Conceição de Paiva,

and Celeste Rodrigues, 258–78. Amsterdam: John Benjamins. https://benjamins.com/catalog/ihll.14.10dua.

Villalva, Alina, and Marta Almeida. 2004. 'Verbos abundantes: usos, desusos e alguns "abusos"'. *XX Encontro da associação portuguesa de linguística*, 14.

Villalva, Alina, and Carlos Alexandre Gonçalves. 2016. 'The Phonology and Morphology of Word Formation'. In *The Handbook of Portuguese Linguistics*, edited by W. Leo Wetzels, João Costa, and Sergio Menuzzi, 1st ed., 167–87. Wiley. https://doi.org/10.1002/9781118791844.ch29.

Vincent, Nigel. 1999. 'The Evolution of C-Structure: Prepositions and PPs from Indo-European to Romance'. *Linguistics* 37 (6): 1111–53.

———. 2001. 'LFG as a Model of Syntactic Change'. In *Time over Matter: Diachronic Perspectives on Morphosyntax*, edited by Miriam Butt and Tracy Holloway King, 1–42. Stanford: CSLI Publications.

Vincent, Nigel, and Kersti Börjars. 2010. 'Complements of Adjectives: A Diachronic Approach'. In *Proceedings of the LFG10 Conference*, edited by Miriam Butt and Tracy Holloway King, 458–78. Stanford: CSLI Publications.

Wagner, Susanne. 2003. 'Gender in English Pronouns: Myth and Reality'. PhD Thesis, Freiburg: Albert Ludwig University.

Wechsler, Stephen. 2011. 'Mixed Agreement, the Person Feature, and the Index/Concord Distinction'. *Natural Language & Linguistic Theory* 29 (4): 999–1031. https://doi.org/10.1007/s11049-011-9149-x.

Wechsler, Stephen, and Hyun-Jong Hahm. 2011. 'Polite Plurals and Adjective Agreement'. *Morphology* 21 (2): 247–81. https://doi.org/10.1007/s11525-010-9168-z.

Wechsler, Stephen, and Larisa Zlatić. 2000. 'A Theory of Agreement and Its Application to Serbo-Croatian'. *Language* 76 (4): 799–832. https://doi.org/10.2307/417200.

———. 2003. *The Many Faces of Agreement*. Stanford Monographs in Linguistics. Stanford: CSLI Publications.

———. 2012. 'The Wrong Two Faces'. *Language* 88 (2): 380–87. https://doi.org/10.1353/lan.2012.0043.

Wescoat, Michael T. 2002. 'On Lexical Sharing'. PhD Thesis, Stanford University.

———. 2005. 'English Nonsyllabic Auxiliary Contractions: An Analysis in LFG with Lexical Sharing'. In *Proceedings of the LFG05 Conference*, edited by Miriam Butt and Tracy Holloway King, 468–86. Stanford: CSLI Publications.

Xavier, M. Francisca, M. de Lourdes Crispim, and Graça Vicente. n.d. 'Dicionário de Verbos Do Português Medieval'. Accessed 16 December 2020. https://cipm.fcsh.unl.pt/gencontent.jsp?id=5.

Yuasa, Etsuyo, and Elaine J. Francis. 2003. 'Categorial Mismatch in a Multi-Modular Theory of Grammar'. In *Mismatch: Form-Function Incongruity and the Architecture of Grammar*, edited by Elaine J. Francis and Laura A. Michaelis, 179–227.

Zaenen, Annie, Joan Maling, and Höskuldur Thráinsson. 1985. 'Case and Grammatical Functions: The Icelandic Passive'. *Natural Language & Linguistic Theory* 3 (4): 441–83.

Zilles, Ana M. S. 2005. 'The Development of a New Pronoun: The Linguistic and Social Embedding of A Gente in Brazilian Portuguese'. *Language Variation and Change* 17 (1): 19–53.