

The Syntax–Prosody Interface in Lexical Functional Grammar

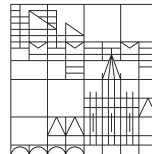
Dissertation submitted for the degree of
Doctor of Philosophy

presented by

Tina Bögel

at the

Universität
Konstanz



Faculty of Humanities
Department of Linguistics

Konstanz, 2015

Abstract

This thesis develops a new approach to the syntax–prosody interface and establishes the integration of the phonological module into Lexical Functional Grammar (LFG).

LFG is a modular grammar theory, which (among other questions) is interested in the relation between FORM and MEANING, i.e., between what is said/perceived and what is intended/understood. An important factor with respect to this question is the distinction between two perspectives that are essential for the communication between speaker and listener: 1) *comprehension*, which discusses the question as to how information from a concrete speech signal influences syntactic phrasing and with it the fundamental ‘understanding’ of what is being said. And 2) *production*, which is concerned with the question how the speaker’s intention is transformed into an utterance. The focus in this thesis is on a specific fragment in this larger model of communication: the syntax–prosody interface.

Given a concrete speech signal, the prosodic grouping of its elements is, on the one hand, related to the language’s internal syntactic structuring: Prosodic phrasing can influence syntactic phrasing in cases of syntactically ambiguous constructions, and syntax also determines prosodic phrasing in that the grouping of prosodic structure to a certain extent reflects (and is thus in parts determined by) syntactic structure.

However, on the other hand, the edges of syntactic and prosodic constituents are frequently incongruent. One reason for this non-isomorphism between syntactic and prosodic phrasing is the rephrasing of prosodically unstressed material. For example, in some languages, function words can be prosodically phrased with a preceding (stressed) element, but are syntactically phrased with a following syntactic head. Such incongruencies between the syntactic and the phonological module can also be found in another group of elements, namely prosodically deficient clitics that can change their position in the clause as a result of specific prosodic requirements. Thus, it can be concluded that prosodic phrasing cannot be solely determined by syntactic phrasing. Instead, processes of prosodic restructuring have to be assumed independently of the syntax–prosody interface. The resulting underlying research question in this thesis is how this tension between intermodular communication and frequent non-isomorphism between syntactic and prosodic structure can be accounted for and how the different components at the syntax–prosody interface have to be arranged to allow for a straightforward analysis of a wide range of language phenomena.

With its modular architecture and intermodular projection functions, LFG constitutes the perfect environment for explorations at the interfaces. For the syntax–prosody interface, this thesis proposes a two-channelled communication between the syntactic and the phonological module: The *transfer of structure*, which abstractly encodes information on how a particular string of language is phrased in each module, and the *transfer of vocabulary*, which exchanges concrete information on lexical properties. In order to account for these transfer processes and for the restructuring of prosodic structure two modifications/extensions to the grammar are made: First, the lexicon as traditionally assumed in LFG is extended to include phonological information as well. Second, a new representation for the phonological module is developed, the p-diagram, which allows for a compact representation of relevant phonological (and prosodic) information. In combination with postlexical phonological processes, this representation constitutes a fully functional and independent phonological module, which can account for a wide variety of postlexical phonological processes including prosodic restructuring (and with it non-isomorphism between syntactic and prosodic structure).

This approach to the interface is supported by the analysis of specific phenomena in a number of different languages/dialects: German, Swabian, Degema, and Pashto. Each of these languages contributes a specific phenomena that highlights (and challenges) a particular aspect of the interface, among them also ‘difficult’ cases like second position en(do)clitics. However, as this thesis shows, all of these phenomena can be accounted for by means of a specific model of the syntax–prosody interface.

Zusammenfassung

Die vorliegende Arbeit befasst sich mit der Schnittstelle von Prosodie und Syntax und mit der Integration des phonologischen Moduls in die Lexikalisch-funktionale Grammatik (LFG).

LFG ist eine modulare Grammatiktheorie, die sich unter anderem für die Frage interessiert, wie die Form einer Aussage und ihre Bedeutung in Verbindung stehen. Aus diesem Grund ist die Frage nach der Art der Kommunikation ein wichtiger Faktor in dieser Arbeit: Handelt es sich bei einer Analyse um das Verstehen von Sprache, wo ein konkretes Sprachsignal in eine semantische Struktur verwandelt werden muss, oder ist sie in einen Akt der Sprachproduktion eingebettet, wo die Intention eines Sprechers in ein Sprachsignal verwandelt werden muss? Der Fokus liegt hier auf einem spezifischen Teil dieser Skala zwischen Form und Bedeutung: der Schnittstelle zwischen Prosodie und Syntax in der LFG, die durch ihre modulare Ausrichtung die perfekte Umgebung für solch eine Untersuchung darstellt.

Einerseits scheinen die beiden Module miteinander in Verbindung zu stehen: Die prosodische Phrasierung in einem Sprachsignal kann im Falle einer zweideutigen Struktur unter Umständen diese Struktur desambiguieren. Die syntaktische Phrasierung wiederum scheint zu einem annähernd großen Teil die prosodische Phrasierung mitzubestimmen.

Andererseits jedoch sind die prosodischen und syntaktischen Konstituenten häufig inkongruent. Ein möglicher Grund dafür ist die Phrasierung von unbetontem Material, das in manchen Sprachen vorausgehendem betontem Material zugeordnet wird, jedoch zur gleichen Zeit syntaktisch mit dem darauffolgenden Wort verbunden ist. Eine weitere Gruppe von Elementen, die zu Inkongruenzen zwischen dem syntaktischen Modul und dem phonologischen Modul führen, sind unbetonte Klitika, die in manchen Sprachen aufgrund von prosodischen Einschränkungen ihre Position im Satz wechseln können. Aus diesem Grund kann die prosodische Phrasierung nicht allein durch die syntaktische Phrasierung bestimmt sein: Eine prosodische Restrukturierung, die nichts mit der Schnittstelle zwischen Prosodie und Syntax gemein hat, kann angenommen werden.

Die zugrundeliegende Frage in dieser Arbeit ist daher die Frage, wie die Spannung zwischen der Schnittstelle als dem strukturellen Kommunikationspunkt zwischen den Modulen und der häufigen Inkongruenz der beiden Module aufgelöst werden kann.

Darüber hinaus stellt sich auch die Frage wie die einzelnen Module und Untermodule angeordnet sein müssen um die Analyse vieler verschiedener Phänomene zu ermöglichen.

Diese Arbeit entwickelt zwei Arten der Kommunikation an der Schnittstelle zwischen Phonologie und Syntax: den Austausch von Struktur, der die im Modul gegebene Phrasierung kodiert, und den Austausch von lexikalischer Information, der konkrete Informationen zu lexikalischen Eigenschaften vermittelt. Um diesen Austausch zu ermöglichen werden zwei Veränderungen an der Grammatiktheorie durchgeführt: Das traditionell morphosyntaktische Lexikon wird um die benötigte phonologische Information erweitert, und eine neue Struktur zur Darstellung des phonologischen Modules wird entwickelt, das P-diagramm. Dieses erlaubt eine kompakte Repräsentation phonologischer und prosodischer Information. Kombiniert mit postlexikalischen phonologischen Prozessen bildet das P-diagramm ein unabhängiges und voll funktionsfähiges Modul, das die Analyse zahlreicher postlexikalischer phonologischer Prozesse ermöglicht.

Um dieses neu entwickelte System zu erproben wurden Sprachphänomene aus verschiedenen Sprachen und Dialekten analysiert: Deutsche Genitiv-Dativ Alternation, Schwäbische Pronomina, Degema Endoklitika und Pashto Wackernagel-Endoklitika, wobei besonders die zuletzt aufgeführten eine besondere Herausforderung für Grammatiktheorien darstellen. Wie die vorliegende Arbeit jedoch zeigt können diese Phänomene unter einer bestimmten modularen Anordnung problemlos analysiert werden.

Acknowledgements

First and foremost I would like to thank my supervisors, Frans Plank and Tracy Holloway King. Frans Plank gave me a lot of freedom to pursue my own interests for which I am deeply thankful. His courses have been an inspiration and his knowledge of the relevant literature seems infinite.

Tracy King determined my career as a PhD student when she offered me an internship at PARC many years ago. Ever since she has been a constant and very reliable presence, whose efficiency and calm friendliness have left a deep impression. In particular, I would like to thank her for her extended comments on previous versions of this thesis and on numerous papers written in the past.

The person who first thought I should do a PhD was Miriam Butt. Over the years, she has been a great academic advisor and has shaped my understanding of linguistics with her passion for language. I am deeply thankful for the years of constant support and guidance and her everlasting patience.

Nicole Dehé kindly met with me several times and commented on various aspects of this work. These discussions always provided me with key insights and were extremely helpful with the decision as to where I needed to set the limits.

I would also like to thank Aditi Lahiri who, over the years, has given me valuable feedback on several occasions. It was her who sparked my interest in phonology, in a memorable lesson on lexical phonology and morphology in 2004. Ever since that class I have been interested in phonology.

Bettina Braun has been a very profound resource for all questions related to lexical access during speech perception and the general discussion of how concrete data feeds into theory. She has furthermore helped me on several occasions with respect to the experiment in Chapter 3.

I would also like to thank the LFG audience who provided me with valuable comments, in particular Ronald Kaplan, and Mary Dalrymple, Louise Mycock and John Lowe, who have challenged my approach to the interface on several occasions and thus forced me to reconsider and adjust my conclusions. I am very thankful for this exchange.

Christoph Schwarze has read and discussed my theories on cliticization and has given me helpful comments. On several occasions (generally accompanied with a game of bridge for relaxation), Astrid Krähenmann provided me with access to data

that is hard to obtain. Achim Kleinmann has been my emergency address for all technical obstacles. Thank you!

Among the group of speakers who have provided me with data, two have particularly supported this work by patiently answering numerous questions: Nafees Ur Rehman for Pashto and Ethelbert E. Kari for Degema. I would also like to thank Bettina Birk and Martin Kappus for their help with the Swabian data.

I am deeply grateful for my ‘second home’, the department of linguistics, which provides a friendly and supporting atmosphere. This thesis would also not have been possible without the help and support of my colleagues (and friends) with whom I shared many steps along the way. Melanie Seiss has made math and daily office routine very easy. Special thanks go to the ‘Urdu group’ Annette Hautli-Janisz, Miriam Butt and Sebastian(Jani) Sulger, with whom I shared many ‘academic’ adventures, fantastic road trips, and Anchor Steam. I could not have done this without you!

Last, but certainly not least, I would like to thank my ‘external friends’, who have stayed with me through the years, in particular Irene, Martina, Kyle, Clemens, Dana, Benni and Iris.

Yet, the greatest and deepest thanks go to all of my family, my parents, who have supported me every step along the way and who, together with my sister and brothers, have helped me uncounted times by taking care of the children when time was short. Amra and Ravi have probably suffered the most from this thesis, but even though they are still very young, they have been extremely patient with me, especially in the last few months.

And without Miriam, this thesis would never have happened.

Abbreviations

In this thesis, the gloss largely follows the *Leipzig Glossing rules*. However, in cases where examples were taken from other authors, the respective gloss was retained. In addition to gloss, the following abbreviations are used in this thesis:

AVM	Attribute-value matrix
DRT	Direct Reference Theory
LFG	Lexical Functional Grammar
LCC	Lexical Category Condition
OT	Optimality Theory
PA	Parallel Architecture
PH	Prosodic Hierarchy
SCB	Serbian-Croatian-Bosnian
SLH	Strict Layer Hypothesis
SPE	The Sound Patterns of English
U	Utterance
ι	intonational phrase
φ	phonological phrase
ω	prosodic word
Σ	foot
σ	syllable

Contents

1	Introduction	1
2	Background	5
2.1	Introduction	5
2.2	Modularity	5
2.3	Introduction to Lexical Functional Grammar	7
2.3.1	C(onstituent)-structure and f(unctional)-structure	8
2.3.2	The lexicon and the <i>principle of lexical integrity</i>	9
2.3.3	Correspondence relations: communication between structures	10
2.3.4	The correspondence architecture	13
2.4	A comparison of architectures	14
2.4.1	The Parallel Architecture	14
2.4.2	The T-model	16
2.5	An introduction to prosodic phrasing	18
2.5.1	The <i>parallel</i> approach	18
2.5.2	The <i>derivative</i> approach	22
2.5.2.1	The Prosodic Hierarchy (PH)	25
2.5.2.2	The prosodic constituents of the PH	26
2.5.2.3	Constraints in the prosodic hierarchy	31
2.5.2.4	Approaches to syntax–prosody mapping	33
2.5.3	A note on (non-)isomorphism	36
2.5.4	Intermediate conclusions	36
2.6	The syntax–prosody interface in LFG	38
2.7	Summary and conclusion	45
3	At the prosody–syntax interface: p-diagram and lexicon	49
3.1	Introduction	49
3.2	The p-structure representation in LFG	50
3.2.1	The prosodic hierarchy: An unstable foundation	53
3.2.2	The representational problem	55
3.3	A syllable-based approach to p-structure	56
3.3.1	The making of the p-diagram	58

3.3.2	A note on pauses	60
3.3.3	The <i>interpretation level</i> of the p-diagram	61
3.4	The multi-dimensional lexicon	64
3.4.1	The lexicon according to Levelt et al. (1999)	65
3.4.2	The lexical model proposed in this thesis	66
3.4.3	The <i>transfer of vocabulary</i>	70
3.5	The <i>transfer of structure</i>	72
3.6	Coordination grouping: an analysis within LFG	74
3.7	Case study: German case ambiguities	77
3.7.1	Experimental set-up	80
3.7.2	Results	81
3.7.3	Discussion and conclusion	84
3.7.4	An analysis of the genitive-dative alternation in LFG	85
3.7.5	Disambiguating c-structure via ranked constraints	86
3.8	Summary and conclusion	89
4	Postlexical phonology and the syntax-prosody interface	93
4.1	Introduction	93
4.1.1	On clitics	94
4.1.2	Postlexical phonological rules	97
4.2	Case study: Swabian pronoun alternation	99
4.2.1	A special case: 1SgNom pronoun subject-drop	100
4.2.2	The prosodic status of 1SgNom pronouns	105
4.2.2.1	The prosodic grouping of ‘function words’ – in general	105
4.2.2.2	The prosodic grouping of Swabian 1SgNom pronouns	107
4.2.3	Postlexical processes in Swabian – an interim summary	109
4.2.4	Swabian pronoun alternation: an LFG analysis	111
4.2.4.1	Information structure status and lexical selection	111
4.2.4.2	From lexicon to p-structure: the <i>transfer of vocabulary</i>	113
4.2.4.3	Postlexical phonology in p-structure	117
4.3	Optimality Theory	120
4.3.1	Optimality Theory in LFG	123
4.3.2	OT applied to the LFG prosody-syntax interface	125
4.4	Summary and conclusion	127
5	Endocclisis at the interface: Degema	131
5.1	Introduction	131
5.1.1	Endocclisis	132
5.1.2	Previous approaches to endocclisis in LFG: Lexical Sharing	135
5.1.2.1	Lexical Sharing and Udi endocclisis	137
5.1.3	String and p-structure in the LFG architecture	139
5.1.3.1	Postlexical phonological rules in an interpreting function	141

5.1.3.2	Postlexical phonological rules in a translation related function	143
5.2	Degema: Some phonological background	145
5.3	The factative clitic	148
5.3.1	Enclisis in the medial position of the intonational phrase . . .	149
5.3.2	En(do)clisis at the final position of the intonational phrase . .	150
5.3.3	Prosodic phrase or syntactic clause	153
5.3.4	Interim summary – paradigm	154
5.4	Degema en(do)clisis: An analysis within LFG	155
5.4.1	Transfer of vocabulary	155
5.4.2	Transfer of structure	156
5.4.3	Postlexical phonological processes in Degema	158
5.4.4	The complete architecture of the syntax-prosody interface . .	160
5.4.5	A note on comprehension and production	162
5.5	Summary and conclusion	164
6	Second Position (endo)clisis: Pashto	167
6.1	Introduction	167
6.1.1	Second position clitics	168
6.1.2	Prosodic inversion: Halpern (1995)	169
6.2	Pashto second position: An overview	170
6.2.1	Syntactic constraints	171
6.2.2	Prosodic constraints	173
6.2.3	Pashto endoclitics	174
6.2.3.1	Monomorphemic verbs: Class I	175
6.2.3.2	Bimorphemic verbs: Class II	175
6.2.3.3	Complex predicates: Class III	177
6.2.3.4	A hybrid: The a-initial verbs	177
6.3	Previous accounts	179
6.3.1	Previous accounts – outside of LFG	179
6.3.2	Previous accounts – in LFG	184
6.4	2P clitics in Pashto: a new proposal	189
6.4.1	The syntactic structure of Pashto	190
6.4.2	The syntax of preverbal clitics	192
6.4.3	An analysis of syntactic 2P clitics in LFG	194
6.4.4	Prosodic second position: the verbal complex	196
6.4.4.1	Preverbal proclitics	197
6.4.4.2	Negative marker	198
6.4.4.3	Interim summary	200
6.4.5	Phonological processes within the verbal complex	201
6.4.5.1	Vowel harmony	202
6.4.5.2	Initial /k/ deletion in complex predicates	205

6.4.5.3	The prosodic phrasing of the verbal complex	206
6.4.5.4	The assignment of phrasing in the verbal complex .	208
6.5	Pashto endoclitics: an LFG analysis	210
6.6	Summary and conclusion	214
7	Summary and conclusion	217
8	Appendix	221

List of Tables

2.1	A short comparison of ALIGN and WRAP.	35
2.2	Lexical entry for ‘university’ (Dalrymple and Mycock 2011).	42
3.1	Lexical entries for <i>Japanese</i> and <i>Amra</i>	69
3.2	Abbreviated lexical entry for <i>übersetzen</i> (‘to translate’/‘to cross over’).	70
3.3	Lexical alignment: the <i>transfer of vocabulary</i>	70
3.4	The German determiner system.	78
3.5	Lexical entries for <i>der</i> and <i>Partner</i>	86
4.1	Postlexical phonological rules/constraints and Swabian clitics.	110
4.2	Part of the lexical entry for the Swabian 1SgNom pronoun.	112
4.3	P-form of the Swabian 1SgNom pronoun.	114
4.4	P-form entries corresponding to example (55).	115
4.5	Phonological rule annotations used in this dissertation	118
4.6	An abstract example for Optimality Theory.	121
5.1	Paradigm for the Tagalog prefix <i>um-</i>	133
5.2	Lexical entries for <i>ùgó</i> ‘vulture’, <i>úgò</i> ‘butterfly’ and <i>úg^lo</i> ‘a kind of stew’.	147
5.3	Endocclisis in Degema	151
5.4	Paradigm of the factative clitic in Degema.	154
5.5	Structure transfer and prosodic rephrasing in cases of non-isomorphism.	163
6.1	Pashto second position clitics (Tegey 1977, 81).	171
6.2	Tegey’s approach: vowel coalescence before 2P clitic placement.	178
6.3	The generation of vowel coalescence according to Kaisse (1985, 142).	180
6.4	Lexical entries as proposed by Lowe (forthcoming, 37, ex. 35) for <i>ʔakwəhə</i>	187
6.5	A selection of possible linear orders of elements within the verbal complex.	201
6.6	Postlexical phonological processes applied to example (125).	205
6.7	Verification of the predictions made by (128) and (128).	208
6.8	Lexical entries for example (130).	210

List of Figures

2.1	C-structure and f-structure representation of <i>Frida sneezed</i>	9
2.2	Association of the lexical information and c-structure.	11
2.3	Relating c-structure and f-structure of <i>Frida sneezed</i>	12
2.4	The parallel projection architecture as represented by Asudeh (2006).	13
2.5	The <i>Parallel Architecture</i> as proposed by Jackendoff (2002).	15
2.6	The language processor (cf. Jackendoff 2002, 197, modified).	16
2.7	The T-model as proposed in Chomsky (1981).	17
2.8	The syntax–phonology interface: SPE and Selkirk (1984).	17
2.9	The Prosodic Hierarchy according to Selkirk (1978).	26
2.10	Lexical and syntactic assignment of ω	30
2.11	The determination of prosodic constituency as assumed in this thesis.	30
2.12	Violations of the Strict Layer Hypothesis.	32
2.13	The end-based approach.	33
2.14	Different mapping approaches for a transitive verb phrase.	36
2.15	P-structure representation in Butt and King (1998, simplified).	39
2.16	The prosody–syntax interface in Mycock (2006, 81).	41
2.17	Prosodically bracketed input to syntactic structure (Bögel et al. 2009).	41
2.18	The string at the heart of the grammar (Dalrymple and Mycock 2011).	42
2.19	The prosody–syntax interface Dalrymple and Mycock (2011, ex.11).	43
2.20	The prosody-syntax interface as proposed in Mycock and Lowe (2013)	44
3.1	C- and p-structure representation of <i>sher e panjaab</i> ‘the lion of Punjab’.	50
3.2	Assignment of tune to ‘ <i>i</i> ’ means ‘insert’ (O’Connor 2004).	51
3.3	Simplified p-structure as proposed by Mycock and Lowe (2013).	52
3.4	The Prosodic Hierarchy	53
3.5	Some phrasing possibilities for <i>John went to school</i>	54
3.6	A possible AVM-representation of [ænə] ‘Anna’.	56
3.7	Speech signal for <i>Ravi</i>	59
3.8	The p-diagram for the acoustic signal for [ravi] (Figure 3.7).	60
3.9	The p-diagram of [ravi]+[pause].	60
3.10	The p-diagram and the P(AUSE)_DURATION dimension.	61
3.11	Speech signal for <i>Ravi and Amra or Karla</i>	63

3.12	P-diagram for the speech signal in Figure 3.11.	63
3.13	Lexical entry for ‘university’ (Dalrymple and Mycock 2011).	64
3.14	Form stratum as proposed by Levelt et al. (1999)	65
3.15	C-structure representations for <i>Ravi and Amra or Karla</i>	71
3.16	Speech signal options for <i>Ravi and Amra or Karla</i>	71
3.17	The prosody–syntax interface in LFG (first version).	73
3.18	The prosody–syntax interface in LFG (second version).	74
3.19	Fragment of the p-diagram for $Ravi \wedge (Amra \vee Karla)$	75
3.20	$Ravi \wedge (Amra \vee Karla)$: at the prosody–syntax interface.	77
3.21	A prototypical <i>dative</i>	82
3.22	A prototypical <i>genitive</i>	83
3.23	P-diagram for the speech signal given in Figure 3.21.	85
4.1	The three-way distinction of clitics proposed in Anderson (2005).	95
4.2	Architecture generally assumed in (post-)lexical phonology.	98
4.3	The position of the lexicon as proposed in this thesis.	99
4.4	Syllabification, subject drop, and invalid syllable structures.	104
4.5	Prosodic grouping of 3SgFem pronouns (Hall 1999).	106
4.6	Existential constraint and inside-out functional uncertainty.	113
4.7	<i>Transfer of vocabulary</i> : metrical, segmental, and prosodic information.	116
4.8	Postlexical phonological rules applied to [vo: =ə =sə].	120
4.9	Two possible integrations of OT at the prosody-syntax interface in LFG.	125
5.1	Two syntactic nodes corresponding to one lexical exponent.	136
5.2	Lexical Sharing tree (Wescoat 2005, ex.22).	137
5.3	Lexical Sharing applied to <i>bey-al-le</i> (watch-FUTII=3SG).	138
5.4	C-structure and p-structure position with respect to the string.	139
5.5	Partial syntactic tree and architecture for s-string = p-string.	141
5.6	Partial syntactic tree and architecture for s-string \neq p-string.	143
5.7	Speech signal for $\acute{u}g^{\downarrow}o$ ‘a kind of stew’ and $\acute{u}g^{\downarrow}o$ ‘vulture’.	146
5.8	p-diagram representations of <i>ugo</i> : ‘stew’ (left) vs. ‘vulture’ (right).	147
5.9	Speech signal for (70): ‘The girl jumped and danced.’	152
5.10	Speech signal for $e=k\acute{o}t\acute{u}=n/_{REL} y\acute{o}$) _i ‘... (they) called DEF’ ((75)).	154
5.11	Arrangement of modules as assumed in this thesis.	155
5.12	Transfer of vocabulary: $\acute{s}m\acute{s} y\acute{o} \acute{s}s^{\downarrow}l$ ‘The child jumped’.	156
5.13	Transfer of structure: $\acute{s}m\acute{s} y\acute{o} \acute{s}s^{\downarrow}l$ ‘The child jumped’.	157
5.14	The micro-components of p-structure.	158
5.15	The preliminary (left) and final p-diagram (right) of example (76).	160
5.16	The complete syntax-prosody interface in LFG.	161
6.1	Postlexical prosodic inversion of a syntactically initial clitic.	169
6.2	Speech signal for (93a) and (93b): $b\acute{x}ylod\acute{a}+me$	176
6.3	Kaisse’s architectural assumption.	180

6.4	The imperfective and the perfective aspect (Roberts 2000, 51, modified).	181
6.5	Syntactic generation of example (97) (Roberts 2000, 58, modified).	182
6.6	Analysis of (102) as proposed by Lowe (forthcoming, 39, (40) (modified)).	188
6.7	C-structure representations for (107a) and (107b).	193
6.8	Abstract representation of the Pashto clause.	194
6.9	Prosodic representations and vowel harmony in examples (121) and (122a).	204
6.10	C- and f-structure representation of <i>wə ye xla</i> ‘Buy it’.	211
6.11	Transfer of structure and vocabulary: <i>ye wəaxle</i> ‘Buy it’.	213
7.1	The prosody–syntax interface in LFG (final version).	219

Chapter 1

Introduction

The aim of this thesis is a precise articulation of the syntax–prosody interface in Lexical Functional Grammar (LFG). While a considerable amount of literature is concerned with syntactic phenomena and a good amount is focussed on the semantic component (and its interaction with syntax), there is comparatively little literature on the syntax–prosody interface and close to none concerned with postlexical phonological processes within LFG. The intention behind this thesis is to establish these aspects of linguistic analysis in the LFG framework, thus clearing the way and providing the tools for the analysis of a much wider range of phenomena.

Among the different phenomena introduced in this thesis, clitics are a very prominent group as they often exhibit a behavior which is challenging to well-established concepts on the syntactic side. However, this is not a thesis on clitic theory, although some of the findings add to the understanding of clitics. But the syntax–prosody interface is associated with a much wider range of open questions ranging all the way up to the architecture of language in general. Addressing all of these is beyond the scope of this thesis.

As such, this thesis is restricted to phenomena that are determined solely by the syntax–prosody interface (although a slight interference of other modules can probably never be excluded). This precludes any phenomena determined by information structure or semantics, e.g., contrastive focus or declarative question intonation. However, a short detour to the relation between information structure and the lexicon will be given in Chapter 4 and the semantic module will appear frequently in the questions concerning the architecture of language as a whole.

Another cut had to be made with respect to the language phenomena under consideration. There are many interesting constructions found in the languages of the world. In the following chapters, four of them will be discussed in more detail: German case ambiguities, Swabian pronoun clitics, Degema en(do)clitics, and Pashto second position en(do)clitics. Each of them was picked because they bring out particular architectural issues.

All chapters in this thesis are written from a specific perspective on the architec-

ture of language. That is, any analysis in this thesis is always positioned between two vanishing points: FORM and MEANING.

MEANING \longleftrightarrow FORM

A directional analysis is either concerned with *comprehension* or *production*. *Production* (or ‘generation’ in computational terms) is concerned with the question of how a speaker *produces* a specific language construction, i.e., how she¹ transforms the original intention, the MEANING, into speech (the FORM). *Comprehension*² describes the ‘other direction’: It is concerned with the question as to how a listener understands what is being said, how he associates the speech with syntactic and semantic constructions. In essence: How does the listener proceed from FORM to MEANING.

At this point, it is necessary to state that the abstract directional representation of language modules and the abstract analysis of a specific phenomenon with respect to these modules and the relations between them should be distinguished from an act of real performance. Instead, it should rather be seen as a ‘road-map’ between the two abstract notions of FORM and MEANING. In a real-time application, this road-map will clearly be employed as well; however, other factors like memory constraints, time, perceptual strategies concerning the identification of lexical elements, frequency of the structure, backtracking and overlapping between modules, and discourse context are, among others, typical issues of ‘performance’ and are *not* in the focus here.³

While the arrangement of the modules between FORM and MEANING sets the bigger picture, the thesis itself is mainly concerned with a fragment of that grammar model, namely the specific analysis of phenomena at the syntax–prosody interface. In the terms of LFG, this interface is between c(onstituent)-structure and p-structure (where “p” represents prosody, (postlexical) phonology and phonetics).

MEANING ... **c-structure** \longleftrightarrow **p-structure** ... FORM

Parts of this thesis build on previously published work. A first approach to the interface as proposed in this thesis has been presented in Bögel (2012), followed by a paper on the genitive-dative alternation in Bögel (2013) and an analysis of Degema endoclefts in Bögel (2014), where each paper incrementally improved the approach to the syntax–interface articulated in this thesis. Pashto endoclefts were discussed in

¹In this thesis, the speaker will always be female and the listener will be male. No gender-related deeper meaning is intended by this choice.

²‘Parsing’ in computational terms; an often used expression is also ‘speech perception’, which is, however often restricted to the association of speech with the lexicon. Thus, the broader term ‘comprehension’ is used here.

³However, concrete speech data in form of an experiment will be analysed and incorporated into the model in Chapter 3. One conclusion that can be drawn from that chapter is also that a clear-cut differentiation between performance factors and abstraction is not easily determined.

Bögel (2010), but the analysis of the data was incomplete and the implementation in terms of LFG was sketchy at most. The Swabian data has never been published.

This thesis is structured as following: Chapter 2 introduces relevant background information: Modularity, language architecture, and different theories of prosodic phrasing are described in addition to an introduction to LFG and a review of previous LFG-related proposals to the syntax–prosody interface.

Chapter 3 introduces different representations of p-structure in LFG and discusses their drawbacks, thus motivating the introduction of a new structure: the p-diagram. This newly established p-structure is then related to c-structure via LFG’s correspondence architecture, thus allowing for information on *structure* to be communicated via the mediation of prosodic constituency. In order to allow for an exchange of lexical information as well, the lexicon is extended to include phonological information. The chapter concludes with an experiment on the dative–genitive alternation in German that allows for the established interface relation to be tested from the perspective of comprehension.

Following this introduction to the newly proposed syntax–prosody interface, Chapter 4 provides a more detailed discussion of the (postlexical) phonological processes applying in p-structure. A case study of Swabian pronoun alternation shows how lexical information is transferred to p-structure, thus focussing on the *transfer of vocabulary*. The following application of a cascade of phonological processes demonstrates how this initial input to p-structure is modified.

Chapter 5 focusses on a very rare phenomena: endocclisis as it is found in Degema. Endocclisis poses a challenge to a lexicalist theory like LFG as it seemingly violates the notion of *lexical integrity* because it alters the string that serves as input to syntax. However, Degema endocclisis can be accounted for by prosodic terms, and after a discussion on the string in general and the resulting conclusion that the phonological (surface) string does not have to be identical to the string as it is processed by syntactic structure, the analysis of Degema endocclisis is straightforward.

In Chapter 6, this challenge is taken one step further, as Pashto second position en(do)clitics require the full force of the proposal made in this thesis. After a detailed (and revised) analysis of Pashto endocclisis, the interface as proposed in this thesis is (successfully) tested against this complex phenomenon. Chapter 7 concludes the thesis

Chapter 2

Background

2.1 Introduction

This chapter introduces crucial background information related to this thesis. The first part is concerned with the architecture of (modular) grammar theories. After a brief introduction into the concept of modularity and a very short excursion into the assumptions made about modules in the field of psycholinguistics, some proposals made by theoretical grammar theories on the architecture of language are introduced.

Next to a more detailed introduction into *Lexical Functional Grammar*, a short introduction will also be given into the assumed architecture of two further grammar theories: 1) *Transformational Grammar* (and its successors) which takes syntax to be the main generative engine of language and is the starting point for most theories of the syntax–prosody interface in the derivative approach (Section 2.5.2), and 2) *Parallel Architecture* as proposed by Jackendoff (2002), which departs from the syntactocentric view and is closely related to psycholinguistic models of speech production/perception as well as to the underlying principles of LFG.

This differentiation between *parallel* and *derivative* is the recurrent theme in this background chapter: Next to the grammar theories, approaches to the syntax–prosody interface also can be roughly associated with either category, although most approaches are a blend of both, with stronger tendencies into one or another direction. The most prominent proposals on the syntax–prosody interface on either side are discussed in Section 2.5. In addition to an introduction to specific topics related to prosodic phrasing, this survey also provides the background for the last part of this introductory chapter: The *derivative* and *parallel* approaches to the syntax–prosody interface in LFG.

2.2 Modularity

The concept of modularity in language is greatly influenced by research into the decomposition of the mind/the brain into faculties, e.g., vision and audition (see

Scheer (2011) for an overview). Fodor (1983) assumes that modules are specialised, high-speed computational systems that process a given input and provide an output according to module-specific vocabulary and principles. Each module is encapsulated/isolated: Once an input has been given, this input is computed in disregard of information coming from other modules.¹ Only when the computation within one module is completed is information transferred to other modules or the ‘central system’ which coordinates the information produced by the different modules.

This modular understanding of the brain has been transferred to the understanding of language as consisting of different modules each with its own set of vocabulary and constraints. The question is in how far these modules are encapsulated and if there is an inherent ‘order’ to them.

Psycholinguistic models of speech production assume the existence of different modules within the ‘language faculty’. In general, each model of speech production shares the main ‘levels of representation’ based on research in speech errors, studies with language-impaired subjects, or chronometric experiments (among others). The commonly assumed levels are the conceptual/semantic, the syntactic, the phonological and the phonetic levels (see Levelt 1999, for an overview). These levels are placed in a specific order depending on either speech perception/comprehension or production. The assumed processing steps given in an act of language production are exemplified by the following quote:

In producing an utterance, we go from some communicative intention to a decision about what information to express – the ‘message’. The message contains one or more concepts for which we have words in our lexicon, and these words have to be retrieved. They have syntactic properties, such as being a noun or a transitive verb, which we use in planning the sentence, that is, in ‘grammatical encoding’. [...] Words also have morphological and phonological properties that we use in preparing their syllabification and prosody, that is in ‘phonological encoding’. Ultimately, we must prepare the articulatory gestures for each of these syllables, words and phrases in the utterance.

(Levelt 1999, 223)

The (ongoing) debate between the different models of speech production is not so much about the existence of the different levels and how they are ordered, but rather on the question in how far these levels overlap during activation. While, e.g., Dell (1986) assumes an interactive cascading model, where activation is spread very far, Bock and Levelt (1994) and Roelofs (1997) assume discrete models, where spreading (and backtracking) is highly constrained.

‘Overlapping’ is also assumed in the opposite direction, in speech perception. As

¹The classical example for such an ‘encapsulated’ module is optic illusion: Even when a subject *knows* about the optical illusion, the vision will still transfer the ‘wrong’ picture. Fodor (1983, 66); an abundance of optical illusion examples can also be found on the internet.

McQueen (2005) in his overview on the process of speech perception shows, activation (and with it the search for an item in the lexicon) starts with the very first sound that a listener perceives. Evidence for overlapping also comes from, e.g., garden-path effects (see also Chapter 3, Section 3.5) which show that readers begin building syntactic structures with the very first word, while eye-tracking experiments can be applied to show that the listener is already exploring possibilities on a screen without having heard the full sentence (see also Jackendoff (2002), Chapter 7 for an exploration into different clues to the overlapping of modules). In short: Modules cannot be viewed as encapsulated units, but output of one module is transferred as input to another module before the overall information flow has been completed.²

As stated in the introduction, the question in how far these modules overlap (and what the resulting consequences in processing are) is not in focus here. Instead, the important conclusion is that modules process a particular part of linguistic information via module-specific principles and constraints and are aligned in a specific way in order to mediate between FORM and MEANING; thus, a possibility to transfer information from one module to the other must be provided (independently from performance factors). The establishment of such a transfer of information between two modules, syntax and prosody, is the aim of this thesis.

2.3 Introduction to Lexical Functional Grammar

LFG is a generative, non-transformational grammar theory, which was developed by Joan Bresnan and Ronald M. Kaplan in the 1970s and 1980s partly in response to Transformational Grammar (Chomsky 1957, 1965, see Section 2.4.2). Based on the statement by (Chomsky 1965, 9) that “this generative grammar does not, in itself, prescribe the character or functioning of a perceptual model or a model of speech-production” and the conclusions made by Fodor et al. (1974) that there is no reliable psychological evidence for transformations (in contrast to structures) and that transformational grammar cannot be employed as a psychologically plausible model, Bresnan (1978) takes it upon herself to develop a grammar theory which includes more ‘psychological reality’.

The first (and well-known) reanalysis in Bresnan (1978) is concerned with passive constructions, which transformational grammar derives from an active counterpart via transformational syntactic movements from a ‘deep’ structure to a ‘surface’ structure. Bresnan (1978) noted that the surface structure resulting from a transformation could also have been generated by the grammar without the assumption of transformation. Bresnan then reanalysed transformations like the active-passive relation as part of an operation on the abstract syntactic argument structure of a *lexical* item. In addition to the emphasis on the lexicon, her approach also eliminates the representation of functional information from phrase structure; instead, the information is stored separately, thus accounting for the generalisation of passive across

²This is, in essence, one possible definition of *parallel* grammar.

languages in functional terms and at the same time representing differences between languages in the surface structure – a first distinction of c(onstituent)-structure and f(unctional) structure.

These preliminary assumptions were followed by a fully-fledged description of LFG in Bresnan and Kaplan (1982), where the name is a composition of the assumptions: *Lexical* refers to the rich and complex structured lexicon assumed by LFG. *Functional*, on the other hand, refers to ‘not configurational’ in that grammatical functions are primitives of the theory and are not determined in terms of phrase structure transformations (cf. Dalrymple 2001).

This original account of LFG assumed two different ways of representing syntactic structure: c(onstituent)-structure and f(unctional)-structure. As these structures and the relation between them are the most-studied ones in LFG, they will be used in the following to explain the underlying principles of LFG and its correspondence architecture.

2.3.1 C(onstituent)-structure and f(unctional)-structure

C- and f-structure are the “two levels of syntactic description” assigned to every sentence of a language (Bresnan and Kaplan 1982, 175). Originally a theory focussed on syntactic phenomena, LFG has grown to be a theory of grammar with ‘levels’ referring to several other types of linguistic information. However, to date, c-structure and f-structure are the most-studied representations. The following introduction into the underlying principles of LFG will therefore concentrate on the relation between these two structures (and the lexicon) before establishing the architecture of the overall grammar as it is commonly assumed in Section 2.3.4.³

C-structure represents the linear and hierarchical organization of words into a syntactic tree. F-structure, on the other hand, encodes the abstract functional organization of a sentence, its predicate-argument structure and the associated functional relations, e.g., subject and object. While c-structure is represented as the traditional syntactic ‘tree’, f-structure is represented as an attribute-value matrix, a set of pairs, where the first member is the attribute and the second member is its value: [ATTRIBUTE value] (in the text represented as ⟨ATTRIBUTE, value⟩). While there is always only one value corresponding to a specific attribute at a specific position in the f-structure (*[ATTRIBUTE value1/value2]), the value can also consist of a nested f-structure: [ATTRIBUTE [ATTRIBUTE value]].

F-structure enables the representation of functional relations, but does not suggest a linear order among them. This property allows for a relatively uniform representation of functional information across different languages.⁴ This separation of

³The LFG analysis of the examples largely follows Dalrymple (2001).

⁴The abstract representation of functional information is the essential motivation for the international *ParGram* group whose aim it is to develop large-scale computational grammars for very different languages with a commitment to parallel f-structure representations (see, among others, Butt et al. (2002), Sulger et al. (2013) for more information). These grammars are implemented

linguistic information and representation is in line with the general notion of modularity: “Each aspect of linguistic structure is organized according to its own cohesive set of rules and principles” (Dalrymple (2001, 85), following Sadock (1991)). In practice this means that LFG does not require different aspects of linguistic information to be of the same formal type; instead, the representation should be determined by the properties of the linguistic information as “Forcing everything into the same mold obscures generalizations and is not conducive to the formation of solid intuitions about the nature and characteristics of the various structures.” (Dalrymple (2001, 89), see also Chapter 3, Section 3.2, and the discussion on p-structure representation).

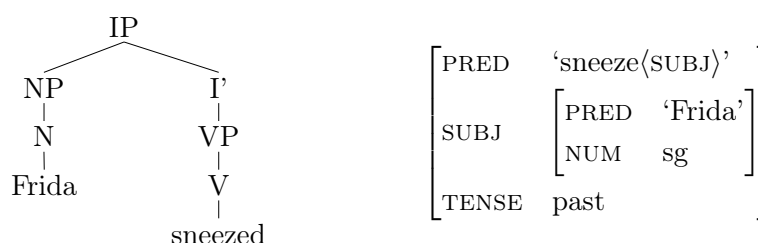


Figure 2.1: C-structure and f-structure representation of *Frida sneezed*.

On the left in Figure 2.1 is the c-structure representation of *Frida sneezed*, encoding linear and hierarchical order.⁵ The f-structure on the right shows the corresponding functional relations. The main predicate *sneeze* is in the top level of the functional structure.⁶ Information which is part of the verb or has functional scope over the clause is stored at the top level as well, in this case information on tense: ⟨TENSE, past⟩. The verb’s subcategorization frame indicates that *sneeze* is an intransitive verb which subcategorizes for a subject. The subject is then encoded as a ‘nested’ f-structure, representing *Frida* as the subject. Further information on *Frida* is encoded as well, namely ⟨NUM, sg⟩. Other possible additions in the subject information could be, e.g., ⟨GEND, fem⟩. This information on the noun *Frida* is stored in the lexicon.

2.3.2 The lexicon and the *principle of lexical integrity*

The lexicon in LFG is a rich and complex structure whose output are fully fledged (i.e., morphologically complete) words. Morphological processes are part of the lexical derivation, i.e., the lexicon does not ‘store’ single affixes that are then merged into

via the grammar development platform XLE (Crouch et al. 2015) and range from small-sized implementations of language-specific phenomena to full-sized, industrially used applications.

⁵Note, however, that the concept of ‘linear order’ will be challenged in this thesis, see Chapter 5/6.

⁶PRED represents the ‘semantic form’, usually represented by the stem. The best way to characterize the words that receive a PRED attribute in LFG is probably with reference to the expression of a ‘content word’ (a word with semantic content). The distinction between lexical and functional categories does not make much sense here, as both are stored in the lexicon in LFG.

the syntactic tree where they are assembled according to syntactic rules. Furthermore, the lexicon is not reduced to lexical exceptions but contains the full range of ‘words’ found in a language. However, LFG does not assume a list of possible forms to be entered into the syntactic tree as needed. Instead, the lexicon is understood as a dynamic component where words are constructed according to internal morpho-phonological processes. As such, the lexical entries as they are represented here are *surface representations* of lexicon-internal complex processes (see also Chapter 3). An example for a lexical entry will be provided in the next section.

LFG is committed to the *strong lexicalist hypothesis*, which states that “no syntactic rule can refer to the elements of morphological structure” (Lapointe 1980, 8). This strict separation of syntax and lexicon is also known as the *principle of lexical integrity* and is one of the basic principles of LFG (see Bresnan and Mchombo (1995) and more recently Asudeh et al. (2013) for discussions of the principle).

(1) **Lexical Integrity:**

Morphologically complete words are leaves of the c-structure tree and each leaf corresponds to one and only one c-structure node.

(Bresnan 2001, 92)

In short, “words are ‘atomic’ at the level of phrasal syntax and phrasal semantics. The words have ‘features’, or properties, but these features have no structure, and the relation of these features to the internal composition of the word cannot be relevant in syntax” (Di Sciullo and Williams 1987, 49). This account of the difference and relation between features and word composition by Di Sciullo and Williams is an exact description of the processes found in the correspondence architecture of LFG, where the features are the relevant information at the interfaces between lexicon, c-structure and f-structure.

2.3.3 Correspondence relations: communication between structures

LFG represents language via parallel structures which encode different linguistic aspects. After the introduction into the principles behind c- and f-structure and the lexicon above, the question remains as to how these modules correspond with each other. This is achieved via correspondence functions (or projection functions) that relate specific parts of one structure to specific parts of another structure; or, formulated in a different way, the structures mutually constrain each other. C-structure and f-structure are related via the projection function ϕ^7 , relating, e.g., the NP containing *Frida* and the f-structure constituted by SUBJ in Figure 2.1.

The constraints that determine the relation ϕ can be added to any node on the right-hand side of a c-structure rule. However, before the correspondence relations can be explained in more detail, the following frequently used symbols have to be explained:

⁷See Dalrymple (2001) for an extended discussion of ϕ .

- $*$ refers to the *current node* in the c-structure tree.
- $\hat{*}$ refers to the *mother node* of the current node in the c-structure tree.
- $\phi(*)$ refers to the *f-structure associated with the current node*.
It is abbreviated by \downarrow .
- $\phi(\hat{*})$ refers to the *f-structure associated with the current node's mother node*.
It is abbreviated by \uparrow .

The following LFG analysis of *Frida sneezed* for the c-structure and f-structure given in Figure 2.1 (and below in Figure 2.3 with annotations) exemplifies the correspondence via projection functions.

The lexical entry:

- (2) *sneezed* V (\uparrow PRED) = ‘sneeze<SUBJ>’
(\uparrow TENSE) = past
- Frida* N (\uparrow PRED) = ‘Frida’
(\uparrow NUM) = sg

The lexical entry for *sneezed* shows the word category (V=verb) which also indicates the position in the c-structure tree. Following is the information that *sneeze* is a predicate and an intransitive verb subcategorizing for a subject. Furthermore, the verb form (i.e., the *-ed* suffix) encodes the information that TENSE = past. This information on subcategorization and tense is made available to *functional structure* via the \uparrow ($\equiv \phi(\hat{*})$) projection. Thus (\uparrow TENSE) = past states that the attribute value pair <TENSE, past> will be part of the f-structure of the lexical entry's mother node in c-structure: V.

Similar, the lexical information on *Frida* will be part of the f-structure of *Frida*'s mother node, which is N. This way, f-structure information associated with a lexical entry is secured to the f-structure associated with the mother nodes in the c-structure tree. That is, lexical items specify their c-structure category and the features they contribute to f-structure in parallel (see also Butt and Kaplan 2002).

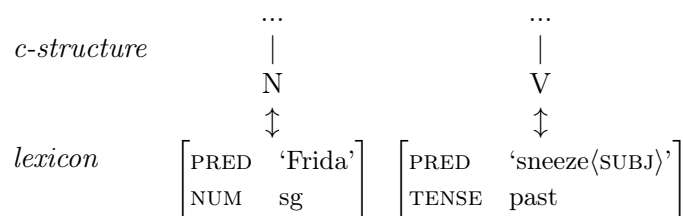


Figure 2.2: Association of the lexical information and c-structure.

However, this is only one part of the information. Other constraints on the f-structure representation for *Frida sneezed* are contributed by c-structure.

C-structure annotation:

$$\begin{array}{rcl}
 (3) \text{ IP} & \longrightarrow & \text{NP} \quad \text{I}' \\
 & & (\uparrow \text{SUBJ}) = \downarrow \quad \uparrow = \downarrow \\
 \text{I}' & \longrightarrow & \text{I} \quad \text{VP} \\
 & & \uparrow = \downarrow \quad \uparrow = \downarrow \\
 \text{VP} & \longrightarrow & \text{V} \\
 & & \uparrow = \downarrow \\
 \text{NP} & \longrightarrow & \text{N} \\
 & & \uparrow = \downarrow
 \end{array}$$

In (3), the specifier of IP, NP, is associated with the subject. The corresponding annotation, $(\uparrow \text{SUBJ}) = \downarrow$, can be read as: “put every f-structure information under my current node ($\downarrow \equiv \phi(*)$) into the value of the attribute SUBJ of my mother node’s f-structure ($\uparrow \equiv \phi(\hat{*})$)”. The second type of annotation given in (3), $\uparrow = \downarrow$, can be read as “put everything of the current node’s f-structure into the mother node’s f-structure”. As this is the default annotation, it is often deleted from a representation.

This results in the following association of c-structure and f-structure for *Frida sneezed* (where the annotations discussed in (3) have been added to the respective c-structure nodes).

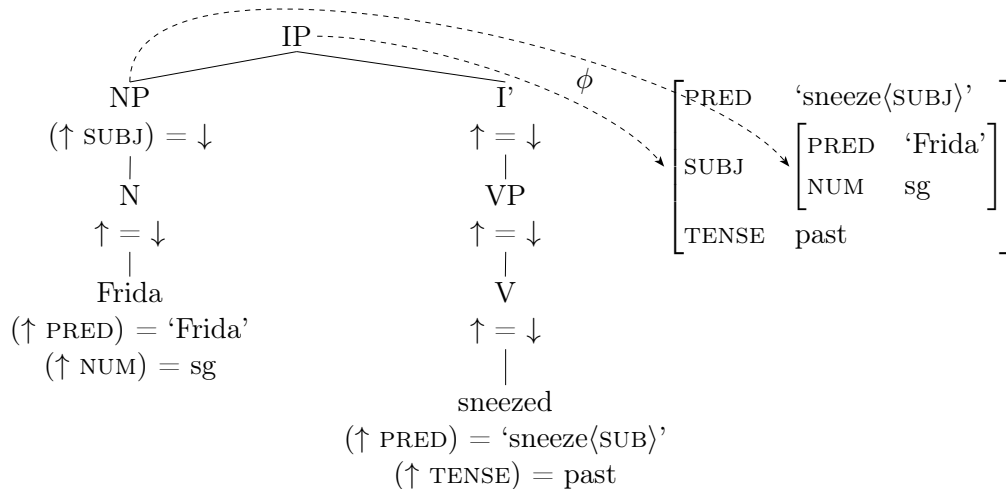


Figure 2.3: Relating c-structure and f-structure of *Frida sneezed*.

Note that it is *not* the case that f-structure is built once c-structure is completed. Instead, f-structure begins to build up as soon as f-structure information ‘becomes available’ with the lexical entry. Thus, in principle, each node in the c-structure

in Figure 2.3 has to be associated with the f-structure on the right, which is not represented for reasons of simplification.

2.3.4 The correspondence architecture

The original architecture of LFG consisted of c-structure and f-structure and the ϕ projection as introduced in the previous section. However, in the last decades, several linguistic components have been added, all of which are subject to their own rules and principles, but all of which are assumed to be present in parallel. The relevant structures for this thesis are the *string*, *c-structure* and *p-structure*.⁸ The most recent assumption of the LFG architecture is represented in the following figure (from Asudeh 2006, 373).

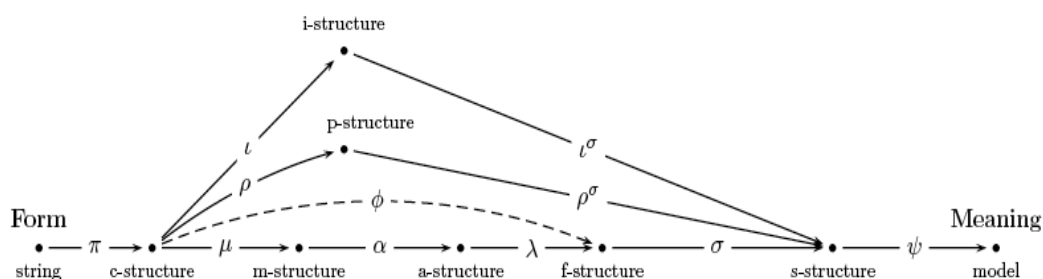


Figure 2.4: The parallel projection architecture as represented by Asudeh (2006).

The projection architecture is located between the two vanishing points FORM and MEANING. The *string* is mapped to *c-structure* via the projection function π and is placed next to FORM, a position that will be revised by this thesis. *P-structure* is projected off *c-structure* via the function ρ as assumed by Butt and King (1998) (see Section 2.6), a position which will also be revised (and was, in fact, revised in Bögel et al. (2009)).

This short introduction into LFG and into the principles of projection functions between structures is superficial at most. Its intention was to provide the reader with the necessary basics and the ‘big picture’ in form of the correspondence architecture relating FORM and MEANING.

Concluding, the following statements can be made about LFG:

1. LFG’s original intention was to model the ‘psychological reality’ of language.
2. LFG is a modular framework; its ‘structures’ represent different levels of linguistic information.

⁸For more information on other structures and relevant references see Dalrymple (2001), Asudeh (2006) and Asudeh and Toivonen (2010).

3. LFG does not assume encapsulated modularity; structures are built up in parallel (overlapping).
4. The different levels of linguistic information are related via projection functions.
5. LFG supports the strong lexicalist hypothesis, the ‘principle of lexical integrity’, which assumes that only fully-formed words enter the syntactic tree.

2.4 A comparison of architectures

There are many grammar theories, most of them with a very specific language architecture. This section will give a brief introduction into the architecture of language assumed in two other grammar theories: *Transformational Grammar* (and its successors) which takes syntax to be the main generative engine of language and thus opposes the ‘psychologically real models’ as introduced in Section 2.2. Interestingly, it nevertheless is the starting point for several popular approaches to the interface, even though the underlying architecture, the T-model, does not provide the optimal input to phonology (or semantics, see also, e.g., Vogel and Kenesei (1990)).

The second grammar theory is Jackendoff’s *Parallel Architecture*, which departs from the syntactocentric view and is closely related to psycholinguistic models of speech production/perception as well as to the underlying principles of LFG.

2.4.1 The Parallel Architecture

Jackendoff’s basic research question is the following: “What is the best way to allocate the generative capacity of language, so as to account for the observed relations between sound and meaning?” (Jackendoff 2010, 584). In contrast to syntactocentric grammar theories, Jackendoff assigns generative capacity to all modules of the grammar. In the Parallel Architecture (hence PA), there are at least three components with their own distinctive primitives that generate their own structure. A set of interface components link the respective structures, thus ultimately establishing a relation between FORM and MEANING.

While Jackendoff explicitly relates to the modularity of language, he also clearly opposes modularity in the sense of Fodor (1983) in that he assumes communication between the modules through a specialized interface. However, this communication is not a derivation of one component from the other; instead, “each structure is licensed by simultaneously applied constraints” (Jackendoff 2010, 588).

One way of communication is the lexical ‘word’, which contains phonological, syntactic, and semantic information. It is thus the perfect candidate for the interface components, as it establishes a correspondence between the three linguistic levels on a lower basis.

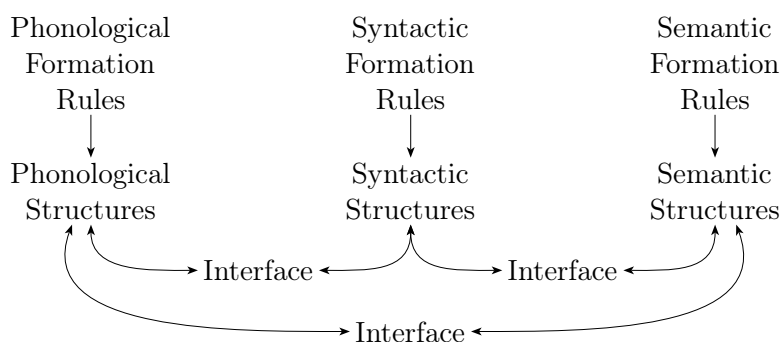


Figure 2.5: The *Parallel Architecture* as proposed by Jackendoff (2002).

So far, PA is very similar to LFG (and vice versa, with the extensions on the lexicon as proposed in this thesis).⁹ However, there is an important conclusion one can draw from PA concerning the tension between a “parallel grammar” and the psycholinguistic models introduced in Section 2.2.

PA is nondirectional, but its constraints can be implemented in any order suited to particular processing tasks.

(Jackendoff 2010, 589)

This states clearly what is only implicitly expressed in LFG in, e.g., Figure 2.4, where the structures assumed in LFG are placed between FORM and MEANING, and what is also reflected in the numerous implementations of LFG grammars (see footnote 4). *Parallelity* refers to the notion of a) generative capacity, in that each component has its own principles and constraints that contribute to the analysis of a language phenomenon, and b) is also a reference to the overlapping of structures (as demonstrated in Section 2.3.3 for the relation between c- and f-structure). It does *not* mean that each component builds a completely isolated structure which then has to be aligned with the output of the other components at some point as proposed by Bögel et al. (2009) and Dalrymple and Mycock (2011) and the follow-up papers (and what essentially is Fodorian modularity). This difference in *parallelity* cannot be stressed clearly enough.

⁹One difference is, for example, the assumption made in PA that affixes are stored separately in the lexicon.

Jackendoff proposes the following model of the PA in language production and comprehension:

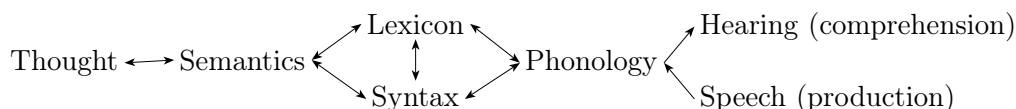


Figure 2.6: The language processor (cf. Jackendoff 2002, 197, modified).

In order to create a processor from a parallel grammar, one needs essentially two ingredients: the structure-internal rules and constraints and the interface constraints that establish a relation between two forms by reducing the communication to the relevant aspects between two structures and by ‘translating’ the information into the language of the ‘receiving’ structure. In LFG, the internal rules are represented by, e.g., c-structure rules while the interstructural communication is encoded in projection functions.

2.4.2 The T-model

As mentioned briefly above, LFG was also developed in response to some of the shortcomings of transformational grammar as first proposed by Chomsky (1957, 1965). As most of the approaches to the syntax–prosody interface discussed in Section 2.5.2 are developed with reference to its underlying architecture (the T-model), this section will give a very brief introduction.¹⁰

Transformational grammar has seen several revisions which usually are accompanied by different names. In its original form, transformational grammar assumed two structures: a *deep structure*, which was associated with semantic relations (and semantic interpretation) and was also assumed to represent properties and relations that were common across languages,¹¹ and a *surface structure*, which closely resembled the phonological representation of a sentence. Surface structure was derived from deep structure via a set of transformations to produce a language’s possible linear surface representations. Surface structure is subsequently readjusted, before a phonetic representation is generated via the application of phonological rules (cf. Chomsky and Halle 1968, SPE).

Chomsky (1981) described a detailed extension (and revision) of this original grammar model, including a *logical form* (LF) and a *phonetic form* (PF) that interact

¹⁰This is, by no means, a complete introduction to transformational grammar (and its revisions and extended revisions) or even to the architectural claims associated with it; the intention is to provide a brief overview on the arrangement of modules assumed in this line of research.

¹¹In that sense, the concept of deep structure and its association with semantics is not so different from the concept of f-structure as assumed in LFG. However, the assumption that surface structure is derived from deep structure by a set of transformational processes/movements that also encode functional relations in surface structure sets the two apart.

with the syntactic component. The resulting grammar architecture is commonly known as the *T-model*.

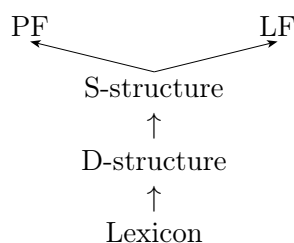


Figure 2.7: The T-model as proposed in Chomsky (1981).

Each component is assumed to include a set of separate rules, where the PF-rules may include some “rearrangement” (Chomsky 1981, 18), and each component is assumed to be blind to the features relevant to another component. While S-structure is mapped independently to both, PF and LF, there is no mapping assumed to take place directly between PF and LF: they are only related through the mediation of S-structure.

Selkirk (1984) adjusts the original account of PF made in Chomsky and Halle (1968) to the T-model: A first set of rules maps from S-structure to PF and ‘translates’ the syntactic representation into a phonological representation, the *rules of the syntax-phonology mapping*. A second set of *phonological rules* derives one phonological representation from another. This account of the syntax-phonology interface is exemplified in Figure 2.8 (modified from Mohanan (1995, 27)).

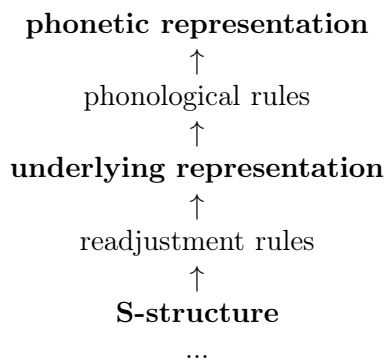


Figure 2.8: The syntax-phonology interface: SPE and Selkirk (1984).

This architectural account of the syntax-phonology interface is the underlying assumption made by most of the interface approaches introduced in Section 2.5.2 with a strong focus on the *readjustment rules*. These include a) the mapping of syntactic structure and b) the readjustment according to phonological principles. The *phonological rules* then operate on the underlying representation, which is, in this thesis

and in most approaches discussed, the constituents of the prosodic hierarchy.

The recent approach to transformational grammar is the *Minimalist Program* which also includes a concept named ‘derivation by phase’ (Chomsky 2001). The main idea behind a phase is that in some point of the syntactic derivation, a ‘phase head’ is introduced which initiates that the completed derivation (part of the material contained underneath the phase-head) is sent off to the semantic and phonological components. As to how exactly the spell-out in phonology takes place is a matter of discussion and as the discussion on the definition of ‘phase head’ is not settled syntax-internally as of yet, the field may change rapidly. See Samuels (2011) for a very good introduction and overview on proposals made with reference to the phonological spell-out.

2.5 An introduction to prosodic phrasing

This section provides a brief introduction into the theory of prosodic phrasing, also including short overviews on the diachronic development of each school of prosodic phrasing, as some of the earlier ideas are recycled in the newer approaches and the difficulties, which have been encountered, have led to various lines of research.

There are two main approaches to prosodic phrasing: One school assumes that prosodic phrasing is relatively independent of syntactic phrasing and is mainly determined by rhythm. From an architectural view, the syntactic and the prosodic module are thus seen as *parallel* in the sense of ‘independent’. The other school believes prosodic constituency to be a *derivative* of syntax in that prosodic phrasing is (mostly) determined by syntactic structure. While the former approach has a very long history, it is nevertheless only occasionally supported in contrast to the latter approach, which has dominated the literature on the relation between syntactic and prosodic structure to a great extent for the last 40 years.

However, as will become clear in the following section, this separation into two schools of thought is not as clear-cut, as the underlying assumptions overlap in various points. Furthermore, the two groups are not homogenous, but a large variety of proposals can be found especially within the derivative approach. Some of the most common ones are discussed below.

2.5.1 The *parallel* approach

One school of thought assumes that phonological constituency is not determined by morphosyntactic constituency; instead, the syllables are grouped together on grounds of rhythmic principles. Such approaches to prosodic structure have a long tradition (see also Plank (2012) for an overview and discussion). Henry Sweet, for example, instructs German learners to correctly pronounce English by providing examples like the following:

- (4) *hij nevə teiksiz hætɔftu eniwan* ('He never takes his hat off to anyone.')
- (Sweet 1904, 74)

Sweet thus groups weak syllables (*his, off, to*) with previous strong syllables forming a trochaic (or dactylic) rhythm to express the prosodic phrasing of the English language to his German readers. He assumes one strong syllable per content word, but he also explicitly distinguishes between stressed content words ('full-words') and unstressed function words ('half-words'), grouping the latter with the former. This results in rhythmic grouping as shown in example (4), where the stressed syllable is always in the initial position of the prosodic chunk, which consequently also results in constructions like the one in (5), where the unstressed initial syllable of *afraid* is grouped together with *I'm*, thus crossing a morphosyntactic boundary.

- (5) *aimə freid* ('I'm afraid')
- (Sweet 1904, 74)

Sweet explicitly notes that this is in contrast to the 'ordinary word division' (Sweet 1876, 3). However, he argues that in English the end of a 'word' is not necessarily discernible except if the sense is known. Thus, if a listener does not know English, he will not be able to distinguish < came tomorrow > from < cameto morrow >. Further support comes from the historical perspective, as this division into sense units in today's orthographic representation is a fairly 'new' phenomenon, after centuries of a literary standard that grouped written words according to their pronunciation into higher prosodic constituents.¹²

Sweet's approach suggests an independence between the two modules of syntax and prosodic constituency. Similar proposals have been pursued by, e.g., Fudge (1999), who discusses the notion that feet must be exhaustively parsed into prosodic words according to the constraints of the prosodic hierarchy (see Section 2.5.2.3) and provides examples where this is not the case (see Fudge (1999) for further information).

Foot structure is also the dominant topic in Lahiri and Plank (2010), who follow Sweet in that they assume the trochaic foot to be the determining element for the creation of prosodic structure, with the stressed syllable as the initial element of each prosodic chunk. They discuss this approach to prosodic phrasing with respect to a number of diachronic and synchronic examples, and an experiment on prosodic grouping in Dutch which supports the assumption that function words are grouped together with preceding strong syllables (Wheeldon and Lahiri 1997).

In the proposal made by Lahiri and Plank (2010), the integrity of the morphological word is not kept intact. Instead, it is possible to group weak initial syllables

¹²In fact, early Greek and Roman scripts are dominated by the *scriptio continua*, a written record of the spoken word which does not distinguish between semantic words, but between larger chunks of speech (see, e.g., Parkes 1992). The habit of grouping words together was also very common in the Germanic languages, as studies of Old English (Baker 2007, among others) and Old High German (Nübling 1992, among others) show. It is only 1100 AD that these traces of phrasing are no longer found in the scripts.

together with strong syllables of the previous word, thus allowing for the crossing of word boundaries. This is exemplified in example (6), which shows the morphological and phonological phrasing of a German saying (see also example (4)):

- (6) **morphological:** Frisch gewagt ist halb gewonnen
phonological: (Frisch ge)(wagt ist) (halb ge)(wonnen)
 with.vigor dared is halfway obtained
 ‘Dared with vigor is halfway obtained.’

(Eisenberg 2006, 136, modified)

In (6), each weak/unstressed syllable is grouped together with a preceding strong/stressed (underlined) syllable, thus forming a trochaic pattern and disregarding morphosyntactic phrasing, which often has a reversed pattern in Germanic languages in that function words usually precede the head.

Note, however, that this approach to prosodic constituency is restricted to the lower levels of the prosodic hierarchy in that the trochaic (dactylic) foot determines the phrasing of the prosodic word; that is, the lower prosodic constituents are determined by rhythmic principles. The major prosodic constituents, on the other hand, are continued to be determined by syntactic phrasing; i.e., the intonational phrase is determined by the boundaries of the syntactic clause (Lahiri and Plank 2010, 374).

Zec and Inkelas (1990) assume a bidirectionality of the phonology-syntax interface at the level of the higher constituents, where each component has a *continuous*, highly constrained access to the other mediated via prosodic constituency. Zec and Inkelas oppose the ‘derivational’ model where one component is positioned after the other, i.e., where prosodic constituency is determined by syntactic constituency. Among other phenomena, they show the necessity for a parallel architecture with an example from Serbian/Croatian/Bosnian (hence SCB) where the topic construction in the syntax is predetermined by prosodic branchingness ((7)).

- (7) a. [[Taj]_ω [čovek]_ω]_{NP} voleo-je Mariju
 that man loved-AUX Mary
 “that man loved Mary”
- b. *[[Petar]_ω]_{NP} voleo-je Mariju
 Peter loved-AUX Mary
 “Peter loved Mary”

(Inkelas and Zec 1995, 545)

According to Inkelas and Zec, only a phonological phrase which consists of more than one prosodic word, can be topicalized in SCB syntax, thus allowing for the construction (7a), but ruling out (7b). Seemingly, this phenomenon requires a parallel

presence of the components, as the syntactic analysis has to know about prosodic constituency and its constraints *before* the decision on syntactic phrasing is made.¹³

The approach presented in this thesis will assume an initially syntactic determination of the higher levels of the prosodic hierarchy (Chapter 5 and 6). Furthermore, it is assumed that prosodic status on the lower levels is assigned as part of the lexical entry (e.g., Chapter 4). However, the approach to the prosody-syntax interface proposed in this thesis allows for a rearrangement of prosodic boundaries as part of postlexical phonology (see Chapter 4). The same mechanism could, in principle, be applied to assign lower prosodic structure on the basis of rhythmic principles ‘from scratch’. However, this possibility is not pursued here for two reasons: 1) the languages discussed in this thesis are not restricted to the Germanic family: Chapter 5 and Chapter 6 will analyse phenomena in Degema (Edoid family) and Pashto (Iranian) respectively. To establish the underlying rhythmic patterns for these languages would have gone far beyond the scope of this thesis. 2) It is not clear how the postlexical grouping of syllables ‘from scratch’ without reference to lexical identity can derive nested structures as they are required for, e.g., Swabian (see Chapter 4). That is, how does an algorithm confronted with a sequence of two syllables, ‘know’ that the phrasing is $((\sigma)_\omega\sigma)_\omega$ and not $(\sigma\sigma)_\omega$.

Nevertheless, it must be noted that the assignment of especially prosodic word status on the basis of morphosyntactic structure is rather controversial, which is briefly discussed in Chapter 3, Section 3.2.1. The difficult assignment of prosodic word status will also be a topic in Chapter 6, where it is indeed the stressed syllable that determines the edge of the prosodic word. The approach presented in this section, where prosodic word status is assigned on the basis of rhythm and not morphosyntactic structure should thus be kept in mind, although some refinements (e.g., on nested structures) have to be discussed and a transfer to other language families would be of interest. The prosody-syntax interface as proposed in this thesis and especially the structure of the postlexical component would be able to support this approach without difficulties.

¹³However, it is difficult to judge from the data if the conclusions are correct, as there are many open questions that would shed more light on these specific constructions and would, possibly, allow for a reanalysis (for example the question why this specific construction is considered to be topicalised and what the non-topicalised construction would look like or if there are any further prosodic indications, e.g., a major break after the topicalised element, etc.). Furthermore, a reanalysis would be possible if considering a linear model and the two possible directions of comprehension and production. From the perspective of comprehension, cases where the topic consists of one prosodic word simply do not exist in real language as they will not be used by a speaker. From the perspective of production, on the other hand, syntax will always allow for two possibilities as output: the topicalised construction and the non-topicalised one. These two options are then expressed in terms of prosodic constituency, which will rule out any topicalised structures with less than two prosodic words in the first phonological phrase. A similar approach was also suggested by Vogel and Kenesei (1990, 351) who proposed a possible overgeneration in syntax to be constrained by prosody.

2.5.2 The *derivative* approach

The derivative approach assumes that prosodic structure is (mostly) determined by syntactic structure. This correlation has been noted very early on by Trubetzkoy (1939/1977, 242) who notes that languages apply *Grenzsignale* ‘boundary signals’ (to a varying degree) to help the listener to parse the incoming speech signal, and to identify sentence, word and morpheme boundaries. Trubetzkoy’s *Grenzsignale* are not materially represented, but they always coincide with morphosyntactic divisions (cf. Scheer (2011, 41)). Pike (1967a) also notes that language can be grouped into levels and comments on their hierarchical organisation: “[...] there are criteria which, in general, differentiate clause level from phrase level, phrase level from word level, word from morpheme level, and so on. [...] unit types on one level must [...] be structurally organized in a manner which in some sense is sharply in contrast with the layer next higher or lower in the hierarchy” (Pike 1967a, 437). Pike lists criteria for each of the levels of this early precursor to the prosodic hierarchy (see Section 2.5.2.1) some of which are strictly phonological but some of which are part of the syntactic component as well; i.e., there is no clear distinction between the syntactic and the prosodic module.

The notion of modularity and the assumption that each module requires a separate set of vocabulary (see Section 2.2) was the start signal for another tradition: the attempt to mark (morphosyntactic) divisions in the speech signal in purely phonetic/phonological terms. In order to distinguish minimal pairs in German, William Moulton (1947) applied so-called ‘juncture phonemes’. A juncture phoneme is represented as /+/, with the following ‘allophones’: /+/, signals a pause at the end of an utterance, but might be realised only as a brief pause within the utterance or alternatively as zero. Moulton also notes that “/+/ occurs almost exclusively at syntactic and morphological boundaries”, but he denies this relationship by assuming that “the phonemes of language should be analyzed without reference to syntax or morphology” (Moulton 1947, 223, fn14). Subsequent literature has defined the notion of juncture in more detail (see Scheer (2012) for an overview); and although the ideas of silent phonemes seems odd from today’s perspective, they are, in a way, the ancestors of the description of suprasegmental units and of the assumption that phonological realities should be expressed in phonological terms and not by reference to other modules of grammar.

An influential approach was *SPE* (Chomsky and Halle 1968, *The Sound Pattern of English*) which assumes that the terminal string produced by syntax consists of segments [+segment] and boundaries [-segment]. Chomsky and Halle note that boundary features do not have phonetic correlates and thus represent the different types of boundaries by a set of symbols: + is a boundary indicating a formative, # is inserted with every lexical category (e.g., N) and with any category which dominates a lexical category (e.g., NP), and = which becomes important in derivation. These different types of boundaries propagated by *SPE* are also representatives of morphosyntactic elements and a mapping results in structures like the following

(shortened from Scheer (2012, 74), cited from Chomsky and Halle (1968, 13)):

$$(8) \quad [_{NP}#[_A#[_N\#tele+graph\#_N] ic\#_A][_N#[_V\#communicate\#_V] ion\#_N]\#_{NP}]$$

$$\quad \quad \quad \downarrow$$

$$\quad \quad \quad \#\#\# telegraph \# ic \#\#\# communicate \# ion \#\#\#$$

SPE also allows for readjustment rules and does not propagate isomorphism between syntactic and phonological phrasing. Chomsky and Halle clearly express this by stating that there are “two concepts of surface structure: input to the phonological component and output of the syntactic component” which show discrepancies that “indicate that the grammar must contain certain rules converting the surface structures generated by the syntactic component into a form appropriate for use by the phonological component” (Chomsky and Halle 1968, 9). The result is that nested syntactic structures are flattened, as, e.g., in the famous example repeated in (9):

$$(9) \quad \text{This is [the cat that caught [the rat that stole [the cheese]]]}$$

$$\quad \quad \quad \text{(This is the cat)(that caught the rat)(that stole the cheese)}$$

(Chomsky and Halle 1968, 372)

Chomsky and Halle note that “intonation breaks precede every occurrence of the category S”, thus allowing for a communication about structure, but they also note that this problem might be part of language performance and does “not belong to grammar – to the theory of competence – at all” (Chomsky and Halle 1968, 372).

SPE was modified in the following years by, e.g., Selkirk (1972) who applies sandhi rules like nasalisation to combinations of lexical and non-lexical categories, establishing amongst other things that the boundary clusters can be reduced to a maximum of two. Furthermore, she clearly states that the external sandhi rules refer to segments and boundaries and not to syntactic phrase structure, thus also establishing modularity.

However, the approach to insert boundaries on the basis of morphosyntactic structure and to use these boundaries during phonological rule application has been strongly opposed by numerous researchers, among them Fudge (1969) who explicitly states that

This framework suffers from the disadvantage of introducing necessary phonological elements in an ad hoc manner, rather than systematically, stating relations between the various elements [...] each element consists of a string of an integral number of instances of the element next below it – a phrase consists of one or more words, a word of one or more syllable

(Fudge 1969, 260)

Pyle (1972) focusses on the fact that the boundaries are claimed to be like segments, but that their behavior is fundamentally different: they have no independent phonological life – they could, in principle, be subject to a set of rules, but these rules never apply in real languages among them.

In reaction to SPE (and its opposition), a number of new approaches were developed in the subsequent years, which can be roughly divided into two groups. McCawley (1968), who strongly opposes SPE, proposes an alternative in form of hierarchically ordered prosodic domains mediating between syntax and prosody, i.e., one of the first clearly expressed versions of the *prosodic hierarchy*. Rotenberg (1978) argues that SPE-boundaries are not part of phonology and should thus not be tolerated. Instead Rotenberg abolishes the translation of morphosyntactic boundaries into SPE-boundaries and replaces them by what can be taken as an early version of the *Direct Reference Theory* which applies relations in syntax as constraints to phonology (see also Elordieta (2008)). The theory of direct reference is briefly introduced in the following, before the following sections will discuss the assumptions behind the prosodic hierarchy.

The main representative of the Direct Reference Theory (DRT) is Kaisse (1985) who defines two relevant factors for the determination of prosodic constituency: c-command, where one words must c-command the other, and the edge condition, where the participants of a phonological rule must lie on the edge of the syntactic constituent that contains them. That is, there is not necessarily isomorphism and uncontrolled reference to syntax: instead, the phonological rules are constrained solely by syntactic relations without reference to the syntactic category of the participants. However, Kaisse's approach has been criticized as not all phenomena can be accounted for by c-command and edge condition (see Elordieta (2008) for an overview). Further disadvantages of the direct syntax approach are, as Hayes (1989) notes, the ability to write phonological rules that are never found in languages (e.g., phonological rules that only apply within a VP). Furthermore, every rule that can be defined as applying within a single prosodic constituent has to be defined for each syntactic possibility, thus making the rules less general.

The DRT approach has not had many supporters, but in recent years researchers (e.g. Seidl 2001, Elordieta 1997) have claimed that some phenomena cannot be analysed in purely phonological terms and with only indirect reference to syntax via the prosodic hierarchy, but that direct access from phonological structure to syntax is needed to some extent. From the perspective of modularity, such claims are difficult as they would require communication between the modules beyond the constituency of the prosodic hierarchy.¹⁴ However, as this thesis assumes modularity *per se* and

¹⁴A close inspection of the data presented by Seidl and Elordieta and a possible reanalysis goes beyond the scope of this thesis and is thus left for further research. Note, however, that another phenomenon discussed in this thesis, the vowel coalescence in Pashto, has been analysed by Kaisse (1985) in the spirit of DRT, but is much easier analysed by means of the prosodic hierarchy, thus preserving modularity (see Chapter 6).

all phenomena presented in the following chapters can be analysed by reference to the constituents of the prosodic hierarchy, the DRT approach is not further pursued here.

2.5.2.1 The Prosodic Hierarchy (PH)

The underlying assumption of the prosodic hierarchy (PH) is the presupposition that spoken language is structured into hierarchically organized prosodic units. These prosodic domains have been approached from various perspectives, the foremost being sandhi processes, whose domain of application can often be defined by a specific prosodic constituent. However, other cues to prosodic constituency have also been observed, e.g., tune association (Beckman and Pierrehumbert 1986, Pierrehumbert and Beckman 1988) and rhythmic grouping (Lieberman 1975, Nespor and Vogel 1989), among others. As Frota (2000, 6) notes, “this means that various types of evidence [...] may be considered in the identification and definition of prosodic domains”.

This possibility to approach the subject of prosodic constituency from different perspectives resulted in (at least) three different versions of the prosodic hierarchy: rule-based, intonation-based, and prominence-based (Frota 2012, 257). The rule-based approach to prosodic constituency focusses on its relation to syntactic structure, but also on prosodic well-formedness conditions (Ghini 1993, among others), while the intonation-based hierarchy refers to the distribution of, e.g., tune and boundary tones. The prominence-based approach, finally, is mainly concerned with rhythmic principles. These different approaches to prosodic constituency are also reflected in this thesis, as the phenomena discussed in Chapter 3 falls into the tradition of intonation-based description, while Chapters 4, 5 and 6 determine the prosodic constituency of the respective language on the basis of phonological processes (rules), and specifically the placement of Pashto en(do)clitics could possibly be argued for in terms of the prominence-based hierarchy.

However, as Frota (2012) notes, there are several indications that allow for the conclusion that there are not three distinct hierarchies, but only one. Hayes and Lahiri (1991), for example, have shown that the prosodic constituents determined by boundary tones in Bengali coincide with the domains of phonological rules. Similar conclusions have been drawn by Frota (2000) for European Portuguese. For this reason and because this thesis focusses on the interface between the phonological and the syntactic component, it will be the *rule-based* prosodic hierarchy that is adopted here. However, as the hierarchy adopted in this thesis also has to account for intonation-based phenomena (Chapter 3), the prosodic constituents are defined more widely (in comparison to definitions made by Selkirk (1978) and Nespor and Vogel (1986)), following the general assumptions about prosodic constituents made by Frota (2012, 261):

A prosodic constituent involves some kind of morphosyntax-to-prosody mapping and an array of phonological properties, including size and prominence,

acting as the domain for phonological and phonetic phenomena (segmental, tonal, temporal), and cues to boundary marking; the morphosyntactic constituent it relates to and at least a subset of the phonetic and phonological properties it shows should be different in type from those defining the other prosodic constituents. In contrast, recursion and compounding refer to forms of grouping of instances of a given prosodic category, yielding levels of phrasing that are reflected only by gradient differences in the strength of the same phonetic properties.

Such a definition is extremely flexible, which also accounts for the fact that it is impossible to find a definition for a prosodic constituent that is justified for all languages, as each language will require an individual set of cues to determine each level of the hierarchy. Nevertheless, in the following, the prosodic hierarchy as it is assumed in this thesis is introduced with a brief introduction as to which cues *can* determine a specific prosodic unit.

2.5.2.2 The prosodic constituents of the PH

Following an earlier proposal by McCawley (1968), Selkirk (1978) introduces a prosodic hierarchy and discusses the motivation for each unit in detail.¹⁵

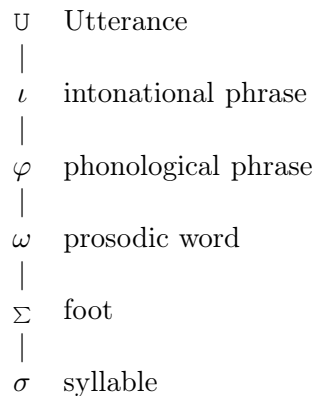


Figure 2.9: The Prosodic Hierarchy according to Selkirk (1978).

In the following, the different levels of the prosodic hierarchy will be briefly introduced with reference to Selkirk and other researchers.¹⁶

¹⁵Note that, in contrast to Selkirk (1972), which supported the insertion of SPE-like boundaries, and Selkirk (1986), which proposes the end-based approach, the hierarchy as presented here was originally also determined by syntactic relations.

¹⁶Note, however, that a complete survey on each constituent is neither given nor attempted, and furthermore, none of the constituents can be considered as ‘settled’ (see also Chapter 3, Section 3.2.1). The following is thus meant as a short overview on recurring/general assumptions.

1. The **syllable** (σ) is taken to be the basic unit of the prosodic hierarchy and is also essential to the p-diagram as introduced in Chapter 3.¹⁷ The importance of the syllable in speech analysis has been discussed from various perspectives. First, as Abercrombie (1967, 34) notes, the perception of the syllable seems to be intuitive in that “Most people seem to be able to say, without much difficulty, how many syllable are contained in a given word or utterance [...] The syllable would appear to be an intuitively recognizable unit [...]”. Abercrombie assigns a physiological nature on the basis of a contraction of the respiratory muscle to the syllable. However, as noted by Ladefoged (1971), this theory is not quite adequate. Nevertheless, Ladefoged also notes that the assumption that articulatory movements can define the syllable cannot be ruled out completely.

Another approach to the syllable relies on its phonotactic structure, which is subject to relatively strict regulations and internal hierarchical principles. The formation of words is thus determined by the constraints on segment alignment within syllables. The string *akpsa* for example could be no word of English as it contains the impossible cluster [kps] (see Kahn (1968, 40); also Fudge (1969)).

Selkirk (1978) discusses the syllable at length, but essentially there are three reasons why she takes the syllable to be the basic entity of the prosodic hierarchy: First it is claimed that a statement on phonotactic constraints of a language can be made by only referring to the syllable. Second, it is only via the syllable that a domain of application for a wide range of segmental phonology can be characterized. And third, it can be argued that suprasegmental processes like stress and tone assignment require a syllable structure, as they are assigned syllablewise (Selkirk 1978, 2).¹⁸

2. Syllables are grouped into **feet** (Σ) which are a metrical unit determined by the rhythmic (i.e., repeated) sequences of stressed and unstressed syllables. These patterns are regular for some languages (e.g., trochaic for German and English, see also Section 2.5.1) and undetermined in others (Catalan, Pashto) (Goedemans and van der Hulst 2013). There are several types of metrical feet, the two most common ones are the trochee ($\sigma\sigma$) and the iamb ($\sigma'\sigma$). Assignment of stress to a specific syllable can be sensitive to syllable weight in which case one speaks of light and heavy syllables (see also footnote 18), but this is not obligatory for every language.

3. Feet are grouped into **prosodic words** (ω).¹⁹ Nespor and Vogel (1986) discuss

¹⁷For this reason, the introduction to this constituent will be more detailed in comparison to the higher units.

¹⁸However, note that some accounts take the *mora* as the basic unit, i.e., “something of which a long syllable consists of two and a short syllable consists of one” (McCawley 1968, 58). Trubetzkoy (1939/1977) notes that there are languages which have the mora as the basic unit and languages which are counting the syllables. The languages discussed in this thesis seem to solely rely on the syllable as the basic unit, but note that the proposal made here can be easily adjusted to include information on morae as well.

¹⁹Prosodic words are occasionally also named ‘phonological words’ (Nespor and Vogel 1986, Levelt

the notion of the prosodic word with respect to several languages and propose that the ω domain is either congruent with the terminal node, or that it consists of a stem, or an element marked by phonological/morphological criteria or the ‘diacritic [+W]’ (Nespor and Vogel 1986, 141). Any ‘unattached’ element in a terminal node form either part of the stem’s ω or a ω themselves. For Nespor and Vogel, a prosodic word is thus never larger than the terminal node.²⁰ A broader definition is provided by Wheeldon (2000, 254) (based on, among others, Levelt (1989)) who defines the prosodic word as “minimally a stressed foot [...] and maximally a single lexical word combined with any associated unstressed function words”. This definition incorporates affixes and clitics into the prosodic word on the one hand, but also allows, e.g., for the prosodic separation of particle verbs in German, where in a word like *abarbeiten* (to carry out successively, e.g., items on a list) the particle *ab-* is *not* included into the prosodic word of *arbeiten*: $(ab)_\omega$ $(arbeiten)_\omega$.

4. Following the prosodic word is the **phonological phrase** (φ) which was defined by Selkirk (1978, 15) as the grouping of the specifier with the head of a syntactic phrase or as the grouping of a function word with its sister constituent. According to Nespor and Vogel (1986), each phonological phrase domain ends with a lexical head. Frota (2012), finally, notes that *syntactically*, phonological phrases relate to syntactic phrases (XP, see also Selkirk (2011)), and *intonationally*, phonological phrases contains a nuclear accent and a phrase accent.²¹

5. The phonological phrase is contained in the **intonational phrase** (ι), which Levelt (1989, 308) describes as “sense unit” corresponding to sentences, clauses and parentheticals. Intonational phrases have a nuclear accent and a boundary tone, as well as a “meaningful pitch contour”. As far the literature is concerned, researchers seem to agree with this categorisation.

6. Not much has been said about the top category of the prosodic hierarchy, the **utterance** (U). Nespor and Vogel (1986, 222) simply define U to consist of all ι s. Hayes (1989, 219) adds that utterances normally relate to a full sentence and that it is framed by pauses, a “cessation of speaking”, but there are also sandhi phenomena related to the edges of utterances (Dehé 2014a).²²

1989). However, as this dissertation will adapt the prosodic hierarchy as proposed by Selkirk (1978), it will also use the term ‘prosodic word’.

²⁰Note, that this is not true for a syntactic theory which assigns separate terminal nodes to, e.g., clitics, as it is the case in LFG.

²¹Beckman and Pierrehumbert (1986) and Selkirk and Tateishi (1988) split the phonological phrase in that the former assume an accentual and an intermediate phrase and the latter a minor and a major phrase. However, this distinction is not relevant in the present thesis, thus, the interested reader is referred to the related literature.

²²Dehé (2014a) demonstrates final devoicing of /l/ at the end of an utterance in Reykjavík Icelandic. Interestingly, while devoicing is always present at the end of an utterance, it is, in decreasing frequency, also present at the right edges of the lower categories. Next to the data

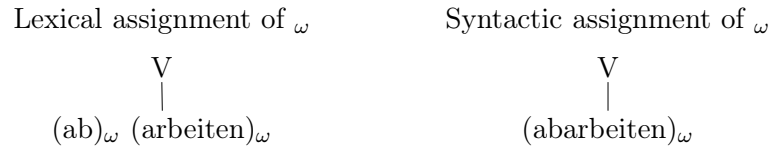
These elements of the prosodic hierarchy as proposed by Selkirk (1978) are commonly assumed in the literature and are also adapted by Nespor and Vogel (1986) and Hayes (1989), with one major difference: they include a further constituent, the **clitic group**, which is situated between the prosodic word and the phonological phrase. Although this constituent is not assumed in this thesis, it is briefly discussed in the following.

Originally proposed by Hayes (1989, (1984))) each content word forms a separate clitic group into which clitics can be incorporated: ((content word)_ω clitic)_C. However, subsequent literature has strongly opposed this notion. Inkelas (1990), for example, provides several arguments against the existence of a clitic group: First, a clitic group cannot be derived on the basis of syntactic terms, as, e.g., *enclitics* may syntactically belong to the preceding syntactic phrase. Second, there is no evidence for a language needing all three constituents: the prosodic word, the clitic group *and* the phonological phrase – the examples provided by Nespor and Vogel can be reanalysed as either applying within the domain of the prosodic word or the phonological phrase. And third, Inkelas also shows that data from Hausa would require the clitic group to be placed *above* the phonological phrase and would thus violate a fundamental principle of the prosodic hierarchy: strict layering (see Section 2.5.2.3). Others have followed Inkelas’s rejection of the clitic group (among others, Lahiri et al. 1990, Selkirk 1995, Meinschaefer 2005) and added further arguments to the discussion (see also Grijzenhout and Kabak (2009, 3) for an overview).

This thesis supports these assumptions and consequently analyses all (clitic) phenomena according to the hierarchy as shown in Figure 3.4, albeit with some modification: Following, among others, Booij (1988), it is assumed that the prosodic structure up to the level of the prosodic word is assigned in the lexicon and not determined by syntactic structure (in contrast to, e.g., Selkirk (1978, 1986, 2011) who projects prosodic word status from the syntactic structure). While the formation of the prosodic word domain in this thesis is further determined during postlexical prosodic restructuring to include clitics, there is also one case where a strict lexical assignment of prosodic word status is difficult, but where a determination of the prosodic word domain solely by postlexical mapping and restructuring constraints is also not a straightforward solution (see Chapter 6). However, cases like the above discussed particle verb *abearbeiten* require the assignment of prosodic word structure in the lexicon if assuming LFG as the underlying framework. As *abearbeiten* will be assigned to a single terminal node, a correct postlexical assignment of prosodic word status will be difficult to monitor and it is for this reason why the prosodic word domain is pre-determined in the lexicon and not mapped from terminal nodes in the syntax-prosody interface.²³

presented in Chapter 3, this is thus another case where *probability* is a factor.

²³Note that the treatment of particle verbs in German is controversial with respect to the lexical/syntactic status, as particles can be syntactically separated from the verbs, as in (10a), but are orthographically part of the verb in (10b) (see also Dehé (2015) and references therein).

Figure 2.10: Lexical and syntactic assignment of ω .

The resulting contribution of syntax and the lexicon to prosodic constituency as assumed in this thesis is represented in Figure 2.11.

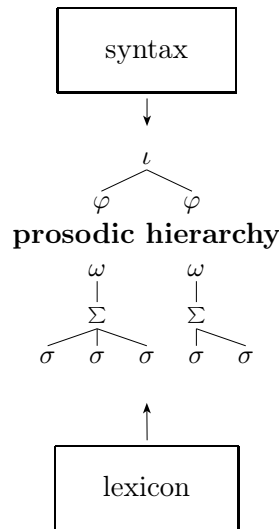


Figure 2.11: The determination of prosodic constituency as assumed in this thesis.

Note, however, that syntactic/lexical structure and prosodic structure are *not* assumed to be isomorphic. This will be discussed in Section 2.5.3 below in more detail.

The prosodic hierarchy as introduced in this section is subject to a number of constraints that regulate the relations among the different levels. These will be

- | | | | | |
|------|----|---|----|---|
| (10) | a. | Wir laden ein
We invite/load PRT
We are inviting! | b. | Wir wollen einladen
We want invite
We want to invite! |
|------|----|---|----|---|

Furthermore, while the meaning of some particle verbs is the composition of its parts, this is not the case for, e.g., *einladen*, as the synchronic meaning of *laden* is ‘load’ (*laden* in the sense of ‘invite’ is rather old-fashioned, although this status might have been acquired very recently in the last 50 years). In their computational analysis of particle verbs in the XLE-LFG grammars, Forst et al. (2010) treat compositional particle verbs via syntactic implementation, but list non-compositional particle verbs in the lexicon. This separation is not adopted here, although it must be stated that no particular theory on particle verbs is assumed: All particle verbs mentioned in this thesis are included in the lexicon in their full form for *reasons of exemplification*: the non-compositional *einladen* as well as the compositional *umfahren* ‘to knock over’ discussed in Chapter 3.

introduced in the following section.

2.5.2.3 Constraints in the prosodic hierarchy

As has been noted by early researchers like Henry Sweet there is a fundamental difference between content words and function words, in that the latter cannot form an independent prosodic constituent. This has been phrased as the *Principle of Categorical Invisibility of Function Words* in Selkirk (1984, 343) and is further developed as the *Lexical Category Condition* (LCC) by Hubert Truckenbrodt.

- (11) *Lexical Category Condition*: Constraints relating syntactic and prosodic categories apply to lexical syntactic elements and their projections, but not to functional elements and their projections, or to empty syntactic elements *and their projections* [highlighted by HT].

(Truckenbrodt 1999, 226)

Apart from the fact that the last part of the LCC does not apply to LFG as there are no ‘empty’ syntactic elements, this is a very important statement as it allows for unstressed elements to be phrased with a preceding stressed element even though they may not share a common immediate XP mother node: thus, non-isomorphism between syntactic and prosodic constituency is given (see also Section 2.5.3).

Although the LCC allows for some flexibility, there are also some restricting constraints that regulate the relations among the different levels that have been applied to the prosodic hierarchy.

(12) Constraints on Prosodic Domination

(where C^n = some prosodic category)

- (i) *Layeredness* No C^i dominates a C^j , $j > i$,
e.g. “No syllable dominates a foot.”
- (ii) *Headedness* Any C^i must dominate a C^{i-1} (except if C^i = syllable),
e.g. “A prosodic word must dominate a foot.”
- (iii) *Exhaustivity* No C^i immediately dominates a constituent C^j , $j < i-1$,
e.g. “No prosodic word immediately dominates a syllable.”
- (iv) *Nonrecursivity* No C^i dominates C^j , $j=i$,
e.g. “No foot dominates a foot.”

(Selkirk 1995, ex. 4)

These constraints are known under the term *Strict Layer Hypothesis* and represent the idea that any level in the prosodic hierarchy can only be composed of the units given in the next level below. However, these strict assumptions soon had to be adjusted. Ito and Mester (1992), for example, proposed a ‘weak layering’ to allow for syllables

to skip the level of the foot and to be directly integrated into the prosodic word. Furthermore, recursivity has been shown to exist for intonational phrases in Ladd (1986) (see also Dehé (2014b)) and all levels below (see van der Hulst (2010) and references therein for an overview). Both violations of the Strict Layer Hypothesis, recursivity and level skipping, are exemplified in Figure 2.12 (Selkirk 2011, 438).

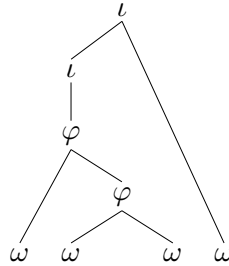


Figure 2.12: Violations of the Strict Layer Hypothesis.

Figure 2.12 contains two violations: a) the intonational phrase ι as well as the phonological phrase φ dominate a prosodic unit of the same category (recursion), and b) ι directly dominates a prosodic word, thus skipping the level of the phonological phrase. These violations are quite common across languages and can also be found in the phenomena discussed in this thesis (especially Chapter 4 and 6).

Further violations include, for example, non-exhaustivity, where feet are not exhaustively parsed under a prosodic word (see Fudge (1999), and also the discussion in Section 2.5.1) and layer violations where the ordering of the domains is not clearly distinguishable. Pak and Frieser (2006), e.g., note that the two phonological rules of accent assignment and liaison in French are applied within prosodic domains that are not strictly ordered with respect to each other, which is a violation of the strict layer hypothesis. They propose several levels of prosodic structuring in PF as a solution, with liaison applying at an early stage and phrase accent assignment at a later stage. Similar accounts can be found in Seidl (2001) and references therein.²⁴

As a result, the decision was made in this thesis to *not* base the p-structure representation on the prosodic hierarchy (see Chapter 3). However, its importance in the determination of a language's phonology and at the prosody-syntax interface is without question which is why prosodic constituency continues to be an important factor in p-structure and has to be accounted for. The following section gives an overview on the most common approaches to the syntax-phonology mapping before discussing the results and conclusions of this Section.

²⁴These cases of layer violations are not considered in this thesis, but are left for further research.

2.5.2.4 Approaches to syntax–prosody mapping

There are several approaches to the mapping from syntactic structure to prosodic structure which relate to the prosodic hierarchy. Four of them will be discussed briefly in this section.

Early approaches to the syntax-prosody interface have related the mapping to specific syntactic information (**relation-based approach**). Selkirk (1978, 15), e.g., forms a phonological phrase by assuming that “an item which is the specifier of a syntactic phrase joins with the head of the phrase” (among other possibilities). Similar relations are defined by Nespor and Vogel (1986), where the phonological phrase “consists of a C[litic group] which contains a lexical head (X) and all Cs on its nonrecursive side up to the C that contains another head outside of the maximal projection of X” (Nespor and Vogel 1986, 168). In a nutshell, the relation-based approach to the mapping from syntax to prosody considers syntactic information such as the relation between specifier and head, head and complement and modifier and head as well as syntactic branching.

However, this implies by no means isomorphism between the syntactic and the prosodic phrase. As Nespor and Vogel (1986, 172ff.) clearly state, the phonological constituent can be restructured, i.e., it can be adjusted independently of syntactic constraints. After the mapping rules have created the prosodic constituents with “reference only to morphological and syntactic structure, the restructuring rules also make reference to the semantic notions and to factors such as the length of the string [...]” (Nespor and Vogel 1986, 300).

In contrast to the consideration of syntactic relations, the **end-based** approach (Selkirk 1986, Chen 1987) is restricted to the edges of heads and maximal projections. The phonological domains (above the foot) are thus assumed to be a stretch of a string demarcated by the left or right edge of a syntactic constituent. The following shortened and modified exemplification is taken from Selkirk (1986, 387) (where ‘fw’ stands for ‘function word’).

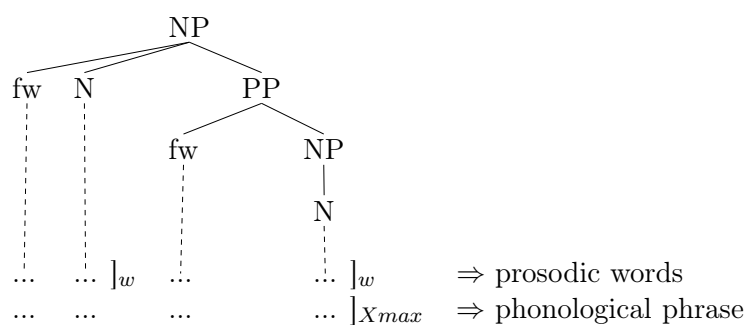


Figure 2.13: The end-based approach.

In Figure 2.13 each content word is marked by a right W-boundary, forming two prosodic words. Each maximum projection is marked by a right Xmax-boundary.

As all right edges of the maximum projections (NP, PP, NP) are congruent, only one phonological phrase is formed. The underlying assumption is also that there are only as many levels underneath the intonational phrase as there are category sets for the end bracket and that the hierarchy of the domains is reflected by the hierarchy among the syntactic categories (cf. Selkirk 1986).

Two parameters have to be considered in this approach to the syntax-prosody interface: the direction of the bracket (to the left or the right of the constituent: ${}_{\alpha}$ or ${}^{\alpha}$) and the parameter setting for α . By applying this simplified version of the syntax-phonology mapping to a number of examples previously discussed in the context of the relation-based approach, Selkirk (1986) shows that her highly constrained (and thus preferable) approach is able to reanalyse critical examples by the mere reference to the ends of constituents.

Note, however, that there is no definite decision made between the two approaches; rather it seems that languages can belong to one or another approach.²⁵ Nevertheless, in recent years, the end-based approach has been very dominant especially in the literature that applies Optimality Theory (OT, Prince and Smolensky (2004), see also Chapter 4) to phonological processes. Within OT, the end-based approach has been reformulated according to the *Generalized Alignment* as formulated by McCarthy and Prince (1993):

- (13) *Generalized Alignment*: Align(Cat1, Edge1, Cat2, Edge2)
 where Edge1 of Cat1 and Edge2 of Cat2 coincide.

Applied to the interface between syntax and prosody, this leads to the following reformulation of the end-based approach to the syntax-prosody mapping (cf. Selkirk 1995):

- (14) ALIGN-XP: ALIGN (XP, R; φ , R)
 which can be translated as “align the right edge of each maximal projection with the right edge of a phonological phrase.”

This group of OT alignment constraints has been extended by another constraint to account for some of the overgeneration caused by the end-based approach: **Wrap-XP** (Truckenbrodt 1999, 228).

- (15) WRAP-XP: Each XP is contained in a phonological phrase.

The difference (and similarities) between WRAP and ALIGN are demonstrated in Table 2.1 (cf. Truckenbrodt 1999, 229).

²⁵Frota (2000, 359ff), for example, shows at length that the phrasing of the European Portuguese data discussed in her thesis cannot be accounted for by the end-based approach.

	A	B
syntactic	$[XP_2 \quad X_1]_{XP_1}$	$[X_1 \quad XP_2]_{XP_1}$
ALIGN-XP,R	$(\quad)_\varphi \quad (\quad)_\varphi$	$(\quad)_\varphi$
WRAP	$(\quad)_\varphi$	$(\quad)_\varphi$

Table 2.1: A short comparison of ALIGN and WRAP.

Option A in Table 2.1 shows an important difference in the two approaches: While ALIGN assigns a right phonological phrase boarder to each XP and ignores the fact that the two XPs are nested, WRAP is able to account for the nested construction by assigning only one phrase boundary, as the separation of mother and daughter XP into two distinct phonological phrases would violate the constraint that each XP is contained in a phonological phrase. This constraint would be violated for XP_1 if the prosodic constituency in option A is determined solely by ALIGN. In option B, on the other hand, the two XPs share the right border and consequently, WRAP and ALIGN make the same predictions. However, the question which of the two results for option A given in Table 2.1 is correct depends on the specific language requirements. Thus, in terms of OT, ALIGN and WRAP are taken to be two competing constraints which have to be ranked anew for each language.

Following the minimalist phase theory (Chomsky 2001), Selkirk (2009, 2011) introduced **match theory** (which has ancestors in, e.g., Ladd (1986)). In contrast to the previous *end-based approach*, *match theory* assumes for both edges of a syntactic constituent to match simultaneously to a prosodic constituent.

- *match clause*: A clause in syntactic constituent structure (S/CP) must be matched by a corresponding intonational phrase (ι) in prosodic constituent structure.
- *match phrase*: A phrase in syntactic constituent structure (XP) must be matched by a corresponding phonological phrase (φ) in prosodic constituent structure.
- *match word*: A word in syntactic constituent structure (X) must be matched by a corresponding prosodic word (ω) in prosodic constituent structure.

(Selkirk 2011, 439, modified)

Match theory reflects the syntactic structure in much more detail in contrast to ALIGN/WRAP while at the same time avoiding reference to specific syntactic relations amongst the members of a constituent. Furthermore, in contrast to the *end-based approach*, *match* predicts a fair amount of recursion, as syntactically nested XPs will be phrased as recursive structures in prosodic constituency as well as the followig example from Selkirk (2011, 464, modified) shows.

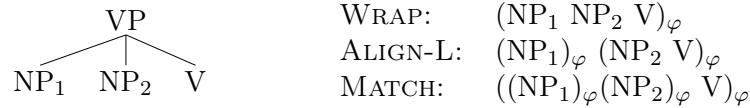


Figure 2.14: Different mapping approaches for a transitive verb phrase.

Figure 2.14 shows that *match* is the only approach that encodes recursive phrasing from the start, mapping each XP into a separate phonological phrase. ALIGN-L(left) on the other hand, only maps the left edges of $\text{NP}_1(=\text{VP})$ and NP_2 , which also includes the following V. WRAP, finally, enforces that every XP has to be wrapped into only one phonological phrase, thus ‘wrapping’ the VP, which also includes both NPs and V.²⁶

2.5.3 A note on (non-)isomorphism

A recurrent topic throughout the last section was the discussion on the (non-)isomorphism between syntactic and prosodic constituents. The common denominator was that it is “crucially [...] not the case that all syntactic boundaries of a certain type must correspond to prosodic boundaries of a given type and vice versa” (Frota 2012, 256). While the *parallel approach* to the interface *per se* expects non-isomorphism at least below the level of the phonological phrase (Section 2.5.1), the *derivative approach* allows for a mapping of syntactic structure to prosodic structure above the level of the foot. However, this does not automatically mean that prosodic constituency reflects syntactic constituency in the derivative approach. As has been mentioned several times in the last sections, most researchers assume “prosodic restructuring”²⁷ (Nespor and Vogel 1986, 172). For example, prosodic phrasing can differ based on the distinction between function and content words (LCC, ex. (11)), or the phonological phrase can be restructured based on size.²⁸

Prosodic constituency is thus not a simple derivative of syntax, but is formed and reformed according to syntax-independent constraints of, e.g., metrical, tonal, information-structural, or lexical origin. Consequently, subcomponents within p-structure that are determined by language-particular requirements can be assumed.

2.5.4 Intermediate conclusions

There are several conclusions that can be drawn from this brief introduction into prosodic phrasing:

²⁶See also Dehé (2014b) for a comparison of these approaches with respect to parentheticals.

²⁷Or *prosodic markedness constraints* in Optimality Theory, see Selkirk (2011, 468ff) for an overview.

²⁸Further factors include the impact of information structure, e.g., focus. This is, however, beyond the scope of this thesis (although a small detour can be found in Chapter 4).

1. The mediator at the syntax–prosody interface is prosodic constituency.

As Zec and Inkelas (1990, 366) note, “Phonological rules will [...] not have direct access to syntactic domains. Likewise, the influence of phonology on syntax will not extend further than prosodic structure, and will at most affect its hierarchical organisation”. This assumption follows directly from the principle of modularity introduced in Section 2.2: Each module has a separate set of vocabulary, i.e., for example, phonological sandhi rules cannot interpret syntactic relations and vice versa. However, communication must be possible to a certain extent as syntactic structure has an impact on prosodic structure and prosodic structure is also a factor for syntactic phrasing decisions (see Chapters 5-6 and Chapter 3, respectively). In this thesis, this communication between the two modules on structural information above the prosodic word is assumed to *only* be based on indirect structural correspondence in form of the prosodic hierarchy.

2. There is no one-to-one mapping between syntax and prosody. Prosodic restructuring will be applied to a preliminary version of prosodic constituency mapped from lexical/syntactic structure.

There is no perfect way of mapping syntactic structure to prosodic structure; that is, any approach to interface mapping can only be an approximation, a preliminary version of prosodic constituency. The final form of prosodic structure, however, is determined by p-structure-internal factors combined with the information provided by other modules (e.g., information-structure). In the following statement made by Selkirk (2011), all vocabulary related to OT has been excluded ([...]).

[...] influences on the phonological domain structure of a sentence are highly modular; it cannot be accounted for by the theory of syntax alone. Rather, a simple theory of the correspondence between syntactic constituency and prosodic constituency posits a set of universal Match correspondence constraints. These interact [...] with phonological constraints [...] to produce a prosodic constituent structure for a sentence which matches up, to greater or lesser degree [...] with the syntactic constituent structure of the sentence. The defining of the phonological domain structure of a sentence is in this sense a true syntax–phonology interface phenomenon, with contributions from the theory of syntactic representation, the theory of phonological representation, and the theory of the correspondence relation between the two.

(Selkirk 2011, 478)

This definition of the syntax–prosody interface could, in principle, include any of the above approaches as it formulates the common denominator: Syntactic influence determines the mapping from syntax to prosody only to a certain extent.

The difference between the approaches is more interesting: While the prosodic word in Lahiri and Plank (2010) is solely determined by foot structure and resulting rhythmic constraints, the derivative approach traditionally assumes the prosodic word structure domain to be (pre-)determined by syntactic heads/content words. Following the brief discussion on particle verbs in Section 2.5.2.1, this thesis assumes prosodic word status to be assigned in the lexicon.

3. P-structure is more than the mere representation of prosodic constituency.

Numerous accounts of literature on phonological phenomena show the necessity to have a separate p-structure representation that allows for a phonological analysis of e.g., rhythmic, tonal, or segmental phenomena and can potentially determine factors like constituent size.

4. The prosodic hierarchy is not a stable foundation.

As discussed in Section 2.5.2.3, the prosodic hierarchy is rather flexible. While the preliminary version mapped from syntax is strictly hierarchical, the hierarchy can include level skipping, recursivity, and possibly even switched layers once it has been modified due to p-structure-internal constraints. Furthermore, as will be briefly discussed in Chapter 5, Section 5.4.5, not every prosodic restructuring that can be identified during a process of speech production is ‘reversible’ during comprehension. It is for this reason, that the prosodic hierarchy is *not* the basic structure of the p-structure representation as proposed in this thesis (see Chapter 3, Section 3.3), albeit its function during the transfer of structure at the interface and its importance in providing structural domains for numerous phonological processes is out of question.

The following section gives an introduction to previous approaches to the syntax–prosody interface in LFG before introducing the new proposal made in this thesis in Chapter 3.

2.6 The syntax–prosody interface in LFG

The literature concerned with the syntax–prosody interface in LFG is comparatively small, but recent years have seen several proposals. This section will give a brief introduction to each approach with a focus on the underlying assumptions and possible problems resulting from these assumptions. Note that there will be a separate introduction to the *p-structure representations* applied by these proposals at the beginning of Chapter 3.

With their analysis of Bengali, **Butt and King (1998)** assume p-structure to represent the interface to a further phonological component where postlexical phonological

processes apply. P-structure itself thus only contains the information that syntactic structure is aware of.

Butt and King assume the prosodic hierarchy as proposed by Selkirk (1978) and propose a direct projection from c-structure, where each clause roughly corresponds to an intonational phrase and each noun phrase to a phonological phrase. Each lexical entry of a content word receives prosodic word status, while clitics are indicated to be part of a prosodic word domain (that must be provided by a content word). Regroupings, e.g., where syntactically unrelated function words are grouped together with a preceding prosodic word are assumed to be resolved between p-structure and the phonological component.²⁹

The representation of p-structure in Butt and King (1998) is an attribute value matrix as shown in Figure 2.15, where the intonational phrase forms the top level of a nested construction which represents the prosodic hierarchy down to the level of the prosodic word.

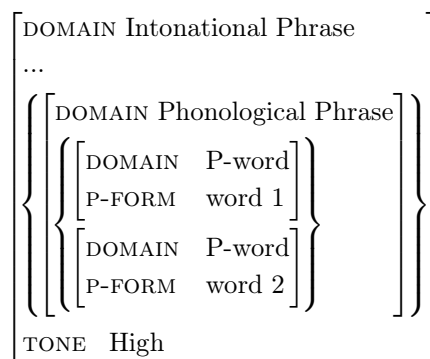


Figure 2.15: P-structure representation in Butt and King (1998, simplified).

For the Bengali data discussed in Butt and King (1998) (which is based on Hayes and Lahiri 1991, among others), not only information on prosodic constituency is projected from c- to p-structure, but also information on tone, in this case indicating neutral/contrastive focus. Thus, depending on the information on focus provided by discourse (d-)structure, the c-structure annotation functions as a mediator between d- and p-structure. A simplified example of this pivot between the two structures is given in (16) for contrastive verb focus, which is indicated by a low tone.

- (16) V \rightarrow ...
 (d::* FOCUS-TYPE) = contrastive
 (p::* TONE) = Low
 ...

²⁹See Bögel et al. (2008) on Urdu Ezafe for another solution to non-isomorphism within this proposal. Note, however, that the computational implementation is very cumbersome.

The current *c-structure* node (indicated by *, here the verb) is marked with the attribute-value pair ⟨FOCUS-TYPE, contrastive⟩ in *d-structure* (d::) and simultaneously projects the attribute-value pair ⟨TONE, Low⟩ to *p-structure* (p::).

However, the assignment of tone to prosodic constituents via projection from *c-structure* is problematic. Bengali assumes that in neutral focus the first prosodic word of the last phonological phrase receives a high tone. This information is again projected from *c-structure* which functions as the mediator between *p-* and *d-structure*. However, the ‘knowledge’ as to which phrase will be the last phonological phrase and which prosodic word will be the target of the high tone is not clear at this point. For this reason, the tone in neutral focus can only be indicated at the level of the intonational phrase and not on the prosodic word itself. This is exemplified in Figure 2.15, where the indicator of the neutral focus, the attribute-value pair ⟨TONE, High⟩, is encoded at the level of the intonational phrase and not in its correct position, namely the prosodic word domain of ‘word 1’. The correct association of tone and target is left to the phonological component.

In the approach presented in this thesis, Bengali tone would be assigned as part of the postlexical component. From a modular perspective, tone assignment is a *p-structure* internal process that cannot be ‘understood’ (or better: ‘handled’) by syntax. As a consequence, the assignment of tone via a projection from *c-structure* is considered to be non-modular.³⁰

Nevertheless, projection of prosodic constituency from *c-* to *p-structure* in Butt and King is very much like the transfer of structure as proposed in this thesis. However, no further phonological component is assumed; instead, prosodic constituency is part of the input to the *p-structure* component along with lexical phonological information (as introduced in the following chapters).

Following Butt and King (1998) is a proposal by **O’Connor (2004)** who implements the mapping between discourse-structure/*i-structure* and prosody in SCB via rewrite rules based on Autosegmental Theory in combination with the ToBI system (Goldsmith 1976, Pierrehumbert 1980). As the focus in O’Connor (2004) is mainly on the *i-structure*/*p-structure* interface it will not be discussed in detail here. Note, however, that a discussion of the *representation* of prosodic information in O’Connor’s approach will be provided in Chapter 3.

In her dissertation, **Mycock (2006)** focusses on constituent questions in several languages and develops a preliminary version of the prosody-syntax interface proposed in Dalrymple and Mycock (2011) below. In this approach, *c-* and *p-structure* are represented as trees; the two modules are interfaced via the lexical items that constitute the string. Each prosodic domain is defined by, e.g., tonal properties or pitch range. In the case of the prosodic word, Mycock assumes it to be correspond-

³⁰An alternative solution, however, cannot be provided at this point as the interface between prosody and discourse (information) structure is explicitly excluded from this thesis. Note, however, that the example on Bengali tone distribution provides evidence for the assumption that it is not (or at least not only) *c-structure* that functions as a mediator between *d-* and *p-structure*, but that some further correspondence between the two modules is likely to exist.

ing to the orthographic word (Mycock 2006, 72), although she notes elsewhere that function words are integrated with the preceding prosodic word (Mycock 2006, 196). How this can be achieved in this proposal is, however, not further discussed. Mycock’s architectural assumptions concerning p- and c-structure are represented in the following figure (which is a fragment of the overall architecture presented in Mycock (2006, 81)).

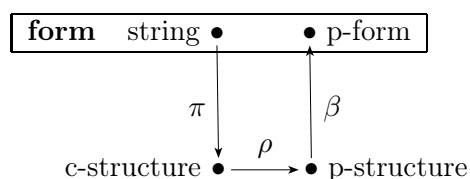


Figure 2.16: The prosody–syntax interface in Mycock (2006, 81).

In Figure 2.16, p-form is the phonological representation of the string (in contrast to p-form as the phonological representation of the lexical element in Butt and King (1998), Dalrymple and Mycock (2011), and subsequent literature). The “string is related to its p-form by virtue of being a projection (via the function β) from its p-structure” (Mycock 2006, 71). Note that, in contrast to the later proposals, this preliminary proposal does not explicitly state that string and p-form have to be parallel.

Bögel et al. (2009) explore the prosody–syntax interface in LFG from the ‘parallel’ perspective, assuming that an independent prosodic component determines the boundaries of prosodic constituents which are then made visible to the syntax in the terminal string.

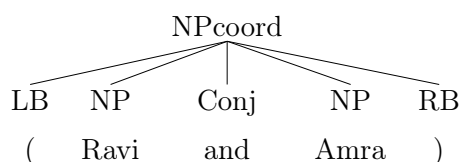


Figure 2.17: Prosodically bracketed input to syntactic structure (Bögel et al. 2009).

Syntactic structures that are incongruent with the prosodic boundaries are dispreferred which makes it possible to account for the disambiguation of syntactically ambiguous structures.

In a follow-up paper, Bögel et al. (2010) discuss second position clitics in Russian and SCB. They develop an account that allows for the clitics to be analysed according to their *functional* contribution in the first position of the sentence in the *syntactic string*, while placing them in second position according to their prosodic requirements in the *prosodic string*, thus defining a mapping between separate prosodic and syntactic strings. These are assumed to be parallel most of the time, but are

allowed to differ in cases like *prosodic inversion* Halpern (1995).

The proposal made in these two papers assumes that the prosodic constituents are formed independently of syntax, which is a different approach from the one proposed in this thesis. Furthermore, the separation of modules is not quite given, as prosodic information is interspersed with the syntactic string.

Albeit not further developed, these two papers set the stage for a very fruitful discussion in the subsequent years and contributed important insights that reemerged in the proposals made by Dalrymple, Mycock and Lowe on the one hand and in this thesis on the other hand.

Following the assumption of parallel independence between syntax and prosody and some of the assumptions made by Mycock (2006), **Dalrymple and Mycock (2011)** develop an account of the prosody–syntax interface where the string is at the heart of the grammar. They develop a lexical entry which contains a s(yntactic)-form and a p(honological)-form, where the s-form represents the (traditional) syntactic information and the p-form encodes the phonological information in form of an IPA representation, syllable structure and syllable weight (Table 2.2).

<i>s-form</i>	university
<i>f-description</i>	N (↑PRED)=‘university’
<i>p-form</i>	/ju _w .ni _w .vəːs.sɪ _w .ti _w /

Table 2.2: Lexical entry for ‘university’ (Dalrymple and Mycock 2011).

Dalrymple and Mycock assume a two-fold representation of the string: A s(yntactic)-string, which serves as input to c-structure, and a p(honological)-string, which is the input to p-structure. The two sides of the string are associated with reference to the lexicon.

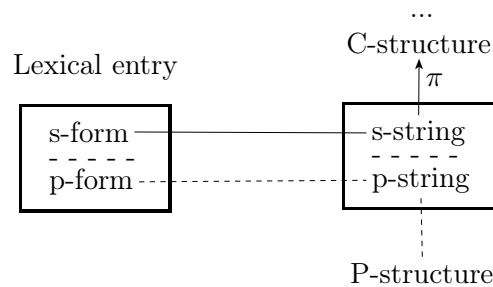


Figure 2.18: The string at the heart of the grammar (Dalrymple and Mycock 2011).

Figure 2.18 shows the two sides of the string at the heart of the grammar associated with the lexicon as a look-up instrument. The dashed lines imply phonological processes: Postlexical phonology is expected to apply between p-string and p-structure formation.

Under this view, c-structure and p-structure are no longer directly related as it

was the case in Butt and King (1998) – the modules are strictly separated. Following Lahiri and Plank (2010), prosodic structure is built on the basis of rhythmic constraints, in that every lexically stressed syllable forms the left edge of a prosodic word. As a consequence, morphological integrity does not have to be preserved. The following figure shows the c-structure and p-structure representation of the casually spoken sentence *Anna was studying at the university* (the default).

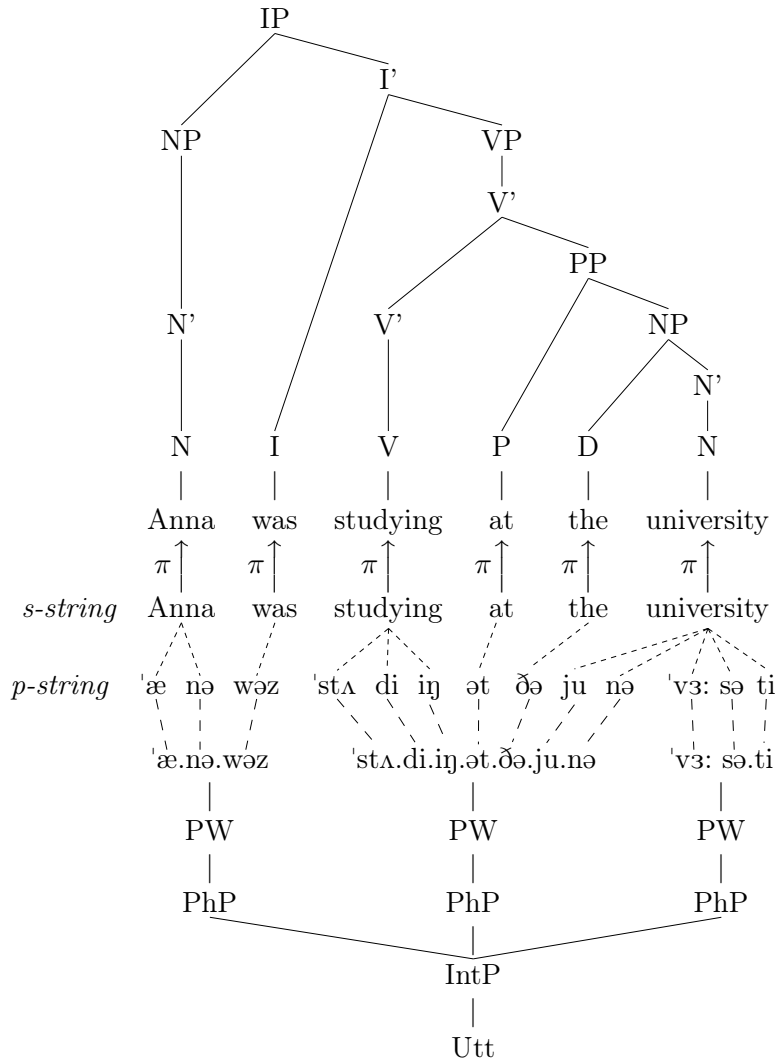


Figure 2.19: The prosody–syntax interface Dalrymple and Mycock (2011, ex.11).

The top half of Figure 2.19 represents c-structure and the bottom half represents p-structure. The interface consists of the s-string and the p-string which are related via the lexical look-up. Note that p-structure is built independently of syntactic principles, solely on the distribution of rhythmic units. To allow communication on

higher levels, information on units of p-structure and c-structure (more specifically: on their edges) are projected into separate structures, e-structure and chi-structure, which are created solely for this purpose. While the default expectation is non-isomorphism between the edges of syntactic and prosodic constituents, Dalrymple and Mycock assume a set of *alignment principles*: In order to communicate information on specific phenomena like declarative question and comma intonation to other structures in the correspondence architecture, Dalrymple and Mycock encode the information on the edges of the respective constituent, where a *principle of interface harmony* ensures that the information on the edge of, e.g., a p-structure constituent, must also be part of the edge in the corresponding c-structure component. How exactly this interface principle is applied and at which point non-isomorphism is assumed, however, is not made explicit.

The architecture proposed by Dalrymple and Mycock (2011) is further developed in **Mycock and Lowe (2013)**, who replace e- and chi-structure projections by assuming that the string should not be represented by atomic elements, but as AVMs encoding lexical information as well as information on the edges of the constituent, which were formerly stored in e-/chi-structure. The resulting modified architecture is shown in Figure 2.20 (a reduced version of Figure 2.19).

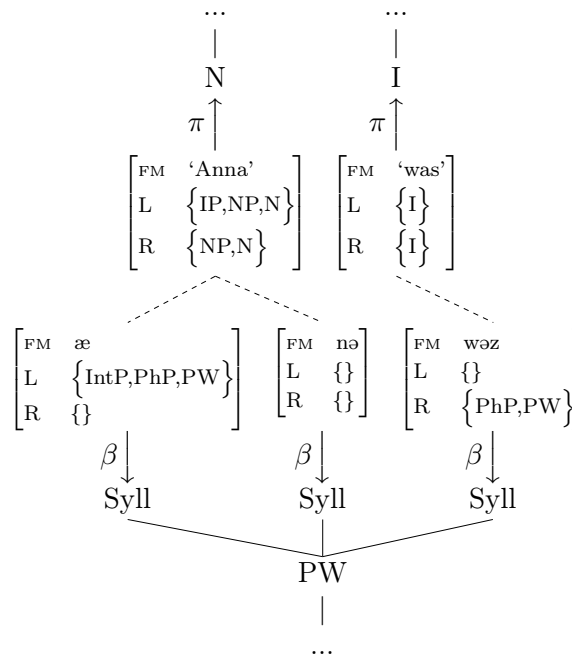


Figure 2.20: The prosody-syntax interface as proposed in Mycock and Lowe (2013)

Instead of single units, the string elements are represented by AVMs. Next to the respective form (FM), these AVMs also encode the position of the element with

respect to the edges of the mother nodes. The syllable /æ/, for example, is at the right edge of the intonational phrase, the phonological phrase and the prosodic word, while its left edge is not shared by any other prosodic constituent (except for the syllable). In this approach, the interface is solely constituted by the alignment of s-string and p-string. If, for example, a phonological element is focussed through a high tone, this information becomes part of the interface as well and, according to the principle of interface harmony, is also part of the other modules.

This strictly parallel approach in the sense that each component is built up independently is seemingly in accord with the “parallel projection” architecture in LFG and strict modularity (in the sense of Fodor (1983)). Note, however, that while the separation of c-structure and p-structure and an assignment of individual vocabulary and principles of formation to each structure are desirable aims (which are also pursued in this thesis), a completely independent construction of prosodic structure is problematic (and also not assumed by any accounts in the literature). It is, for example, not clear how the gap between prosodic word status on the basis of a stressed syllable on the one hand and the phonological phrase on the other hand is closed in this approach, as the phonological phrase cannot be determined solely by rhythmic principles.

As a consequence, a careful distinction must be made between the notion *parallel* in the sense of *independent* and *parallel* in the sense of *building up as soon as relevant information is available, even though the other module is not completed as of yet*. The former follows a proposal made by Lahiri and Plank (2010) *without* sharing their assumption that prosodic structure above the prosodic word is highly influenced by syntactic structure. The latter understanding is essentially what is assumed by the correspondence between c- and f-structure (and the principle of the correspondence architecture in general): *parallel* in the sense that f-structure begins to build up as soon as c-structure is built.

The difference between these two approaches can be seen in Chapter 6 when comparing the proposal made for Pashto endoclysis in this thesis and the analysis provided by Lowe (forthcoming), where it becomes clear that the parallelity between the strings, as it is essential for the framework proposed by Dalrymple, Mycock and Lowe, is untenable for cases where the surface representation is modified by solely *prosodic* requirements.

2.7 Summary and conclusion

This chapter has introduced a wide variety of relevant background information. After discussing the concept of modularity also in the context of psycholinguistic models of speech production and comprehension, a comparison of (theoretical) grammar theories has shown different language architectures, differentiating between parallel architectures and syntactocentric approaches to grammar. In the second part of this chapter, different approaches to the syntax-prosody interface have been introduced,

roughly differentiating between parallel and derivative approaches, although these concepts overlap to a certain extent. The last part presented proposals to the syntax–prosody interface in LFG.

The following conclusions can be drawn from this overview: 1) There is no perfect mapping from syntax to prosody; and 2) all current approaches to the interface are recycled versions of earlier proposals with additional material added, and are to some extent always constrained by the underlying principles of the embedding grammar architecture. Having thus stated the obvious, the question remains as to which conclusions can be drawn for the development of a syntax–prosody interface in LFG. These are given in the following.

Modularity:

Language consists of modules, each with their own primitives and principles. These modules are related via functions that ‘translate’ information into the module’s ‘native’ vocabulary: projection functions in LFG-terms. Modules are not fully encapsulated units, but are built up in ‘parallel’: i.e., fragmental information from one module becomes available to another module upon formation.

Language architecture:

Following the original aim of LFG to develop a language model with ‘psychological reality’, modules are arranged in a specific order to mediate between FORM and MEANING. The respective arrangement of modules, i.e., the direction between FORM and MEANING, is described as *production* and *comprehension* in this thesis. A directional analysis should be understood as a road map, free from questions of performance like ‘working memory’, ‘backtracking’ and ‘overlapping’, although any model must, of course, allow for a performance model to compute; i.e., it must have ‘psychological reality’.

The syntax–prosody interface:

Prosodic constituency is the mediator at the interface between syntactic and prosodic structure. However, next to the transfer of structure, the basic vocabulary for each module must also be provided: This will be achieved by the introduction of a multi-dimensional lexicon in the following chapter.

As has been shown in the previous section, previous accounts of p-structure in LFG do not account for, e.g., the readjustment of prosodic constituency or a complete postlexical phonological analysis. Furthermore, the representations of p-structure proposed so far cannot reflect the full range of prosodic and phonological requirements. For this reason, a new p-structure representation was developed, which will be introduced in the following chapter.

Chapter 3

At the prosody–syntax interface: p-diagram and lexicon

3.1 Introduction

This chapter will introduce the main contributors to the prosody–syntax interface as it is proposed in this thesis, namely the p-diagram as a new representation of p-structure and the multi-dimensional lexicon. Furthermore, this chapter focusses on the underlying architecture of the interface, in that it establishes the correspondence between c-structure and p-structure on the basis of two processes: the *transfer of structure* and the *transfer of vocabulary*.¹

As mentioned in the introduction, any linguistic phenomenon in this thesis is analysed from a particular perspective: from FORM to MEANING (=comprehension) or from MEANING to FORM (=production). While the following three chapters focus on the perspective of production, the current chapter describes the interface from the viewpoint of comprehension. The reason for this 3:1 imbalance lies in the nature of the interface: The syntactic structures that can be used to determine which prosodic cues have a direct impact on syntactic structure are limited to constructions that show a full ambiguity. That is, while there is interesting psycholinguistic research in, e.g., the interaction of prosody and syntax in *garden-path* sentences,² this research is only remotely relevant for the establishment of the prosody-syntax interface as it is pursued here, because the disambiguation in garden path constructions is *syntactic* and not prosodic. With syntactically fully ambiguous sentences, on the other hand,

¹Each of these processes will be discussed in more depth in the following chapters, albeit from the perspective of production. The terminology of *structure* and *vocabulary* has been taken from Scheer (2011, 558). The contexts in which these terms are used in this thesis cover similar processes as originally intended by Scheer (and generally applied in interface theories). Note however, that the implementation here is very different from the one proposed by Scheer (2008, 2012).

²Garden-path sentences are grammatical sentences that have an initial ambiguity which is resolved by the following parts of the sentence. The reader is misled to an incorrect interpretation, which has to be corrected afterwards (see also Section 3.5).

the disambiguation by syntactic means is excluded and the relevance of prosodic cues can be established.

After an introduction to previously proposed p-structure representations in LFG (Section 3.2), two phenomena will be discussed in this chapter: constituent grouping in (English) coordination structures and the genitive/dative alternation in German. The former serves as a demonstration example while the relevant components at the prosody–syntax interface are introduced and discussed step by step: the p-diagram (Section 3.3), the multi-dimensional lexicon (Section 3.4) and the transfer of structure (Section 3.5). These introductory sections are followed by a large-scale production experiment on the dative/genitive alternation in German (Section 3.7), which will also discuss some of the issues related to the integration of spoken data to the prosody–syntax interface.

3.2 The p-structure representation in LFG

Within LFG, only few approaches to encoding prosodic information have been developed. While Chapter 2, Section 2.6 provided an overview on the general assumptions behind these approaches, this section specifically focusses on the way prosodic structure is represented.

Butt and King (1998), analysing N/V-V constructions in Bengali, encode prosodic structure in an attribute value matrix (AVM) which is projected off c-structure. This approach was also applied by Bögel et al. (2008), who analysed ezafe-constructions expressing possessive relations in Urdu. The following Figure 3.1 shows the c-structure and the corresponding p-structure of the ezafe construction in (17).

- (17) *sher=e panjaab*
lion=Ez Punjab
 ‘a/the lion of Punjab’

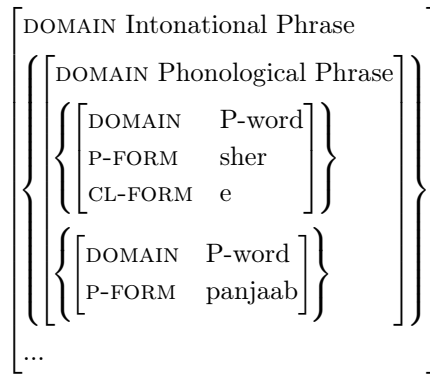
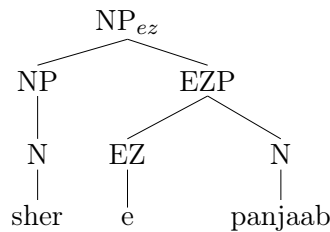


Figure 3.1: C- and p-structure representation of *sher e panjaab* ‘the lion of Punjab’.

Note that while the ezafe clitic syntactically licenses the modifier *panjaab* ‘Punjab’, the clitic is prosodically incorporated into the prosodic word domain of the head

noun *sher* ‘lion’ in the attribute value matrix shown in Figure 3.1. Thus, even though p-structure is directly projected off c-structure, non-isomorphism is possible to a certain extent with this approach. Furthermore, more information can be added within each domain, e.g., information on boundary tones etc. However, the reduction to c-structure as the determining factor of prosodic structure and the restriction of p-structure as a means of prosodic representation neither allows for a concrete representation of postlexical phonological processes nor for an analysis of problematic cases of cliticization, e.g., endoclysis as it will be described in Chapters 5 and 6.

O’Connor (2004) combines the higher units (intonational phrase and intermediate phrase in his account) of the prosodic hierarchy with a description of the speech signal’s fundamental frequency by means of the ToBI framework.³ In his analysis, O’Connor focusses on the interaction between *intonation* and information structure, which is also reflected by his approach as it does not allow for any further information on prosodic/phonological structure to be inserted (even though O’Connor notes that “p-structure represents, in theory, the whole spectrum of prosodic phenomena” (O’Connor 2004, 125)). In contrast to the (broader) AVM-approach of Butt and King (1998), which is, in principle, able to include information from the complete spectrum of the speech signal, O’Connor represents intonation in a tree-like structure, connected loosely with the string. An example is given in Figure 3.2, where O’Connor (2004, 158) assigns the following tune-structure to the sentence ‘*i*’ means ‘*insert*’ (based on an analysis by Beckman and Ayers Elam (1997)).

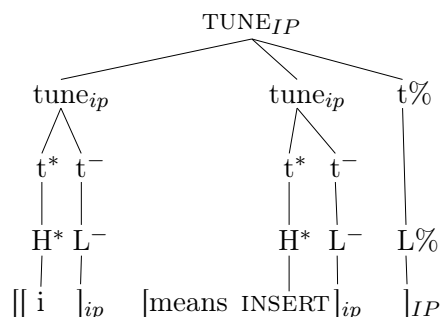


Figure 3.2: Assignment of tune to ‘*i*’ means ‘*insert*’ (O’Connor 2004).

As has been indicated above, the reduction of p-structure to intonation is too narrow as it cannot do justice to the broad spectrum of information found in a sequence of spoken language, neither on the segmental nor on the supra-segmental level. The approach to p-structure presented in this thesis, on the other hand, can account for all of this information and furthermore provides a module-internal structure that allows for the inclusion of postlexical phonological processes as well.

³ToBI represents conventions for assigning High and Low tones to the fundamental frequency of a speech signal, thus describing relevant aspects of the intonational contour (Silverman et al. 1992, see also Section 3.3.3).

Dalrymple and Mycock (2011) develop a prosodic representation to account for comma intonation and question intonation. They place the string (i.e., its two representations, the p-string and the s-string) at the heart of the projection architecture and use it to align a prosodic tree with the c-structure tree. In addition, they project two further structures, which contain the bracketing information relevant for the alignment of the syntactic and the prosodic constituents, information on the intonational contour, etc. In a follow-up paper, Mycock and Lowe (2013) replace the additional structures and propose that s- and p-string should not be represented as atomic elements, but as AVMs. The AVMs representing the p-string are assigned syllablewise and include information on the prosodic constituency (in the sense of right and left boundaries) and other relevant prosodic information, including information on tone.

The resulting p-structure is a combination of a tree-like representation of prosodic constituency with AVMs to encode the additional information in p-structure, e.g., lexical stress or pitch accent. A simplified version of this approach to p-structure is given in Figure 3.3, encoding an intonational phrase starting with [ænə] ‘Anna’.⁴

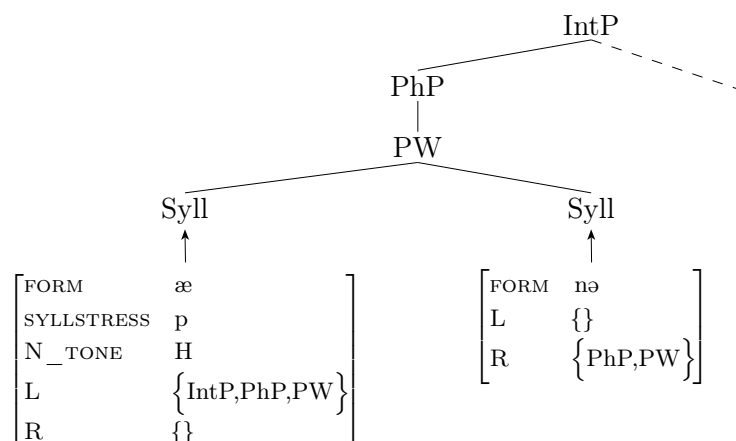


Figure 3.3: Simplified p-structure as proposed by Mycock and Lowe (2013).

To summarize, while the proposals made by Butt and King (1998) and Mycock and Lowe (2013) can in principle include more information, O’Connor (2004) is too narrow for a wider range of prosodic phenomena and is thus discarded. Although the former two differ in their representation and their assumptions about the underlying architecture, they have the same skeletal structure: the prosodic hierarchy. The following section discusses this role of the prosodic hierarchy as the underlying structure for prosodic representation in more detail.

⁴See Chapter 2, Section 2.6 for a more detailed introduction and a complete overview on the prosody–syntax interface proposal by Dalrymple and Mycock (2011) and Mycock and Lowe (2013).

3.2.1 The prosodic hierarchy: An unstable foundation

Most accounts of the prosody-syntax interface are essentially determined by (some version of) the prosodic hierarchy (Selkirk 1978), and consequently, most representations of p-structure have been based on some related hierarchical design. As discussed in Chapter 2, Section 2.5.2.1, the prosodic hierarchy consists of hierarchically ordered prosodic units as shown in Figure 3.4.

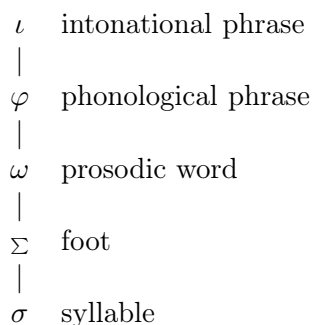


Figure 3.4: The Prosodic Hierarchy

From the perspective of comprehension, these abstract descriptions of prosodic constituents are representations of events in the speech signal: They are indications of processes at the segmental level (e.g., they confine specific segmental-phonological processes) and at the suprasegmental level (e.g., they are indicated by a specific pattern in the pitch and their boundaries might be indicated by a higher duration of the previous syllable, pauses, or an abrupt switch in the pitch). Thus, the elements of the prosodic hierarchy indicate the grouping of smaller and larger groups of segments, and are considered to be the mediator of prosodic structure at the interface to syntax (and possibly other modules), as they represent information on prosodic constituency that could be relevant for a syntactic interpretation (see Section 3.5).

The literature on the question of how these units should be defined and which information from the speech signal should be used for their calculation is huge and quite controversial. For example, a majority of the researchers takes the prosodic word to be *minimally* of the same size as the morphosyntactic stem or even the syntactic word, including affixes. Furthermore, prosodically deficient items (clitics/function words) are assumed to be included into the prosodic word as well (among others, Nespor and Vogel 1986). On the other hand, there are also researchers who propose that a prosodic word starts with the stressed syllable of a trochaic foot (Lahiri and Plank 2010), which in principle allows the morphosyntactic stem to be split by a prosodic word boundary (Chapter 2, Section 2.5.1). In this approach, the prosodic word formation is completely independent from the morphosyntactic representation. Similar discussions exist for other levels of the prosodic hierarchy which makes it very difficult to establish reliable phrase structure rules for prosodic phrasing (in contrast to syntactic phrasing).

Another controversial discussion includes two early concepts of prosodic constituency, *non-recursivity* and the *strict layer hypothesis* (Selkirk 1984, 1986, Nespor and Vogel 1986, see also Chapter 2, Section 2.5.2.3). The body of literature rejecting (or weakening) these concepts, attesting recursion and the possibility of ignoring levels, is quite extensive (see, among others, Seidl 2001, Truckenbrodt 2007, Mester and Ito 2009, Fudge 1999, and references therein as well as Chapter 2, Section 2.5.2.3). Thus, a strict layering of prosodic constituents cannot be assumed, which is another difficulty for the attempt to establish reliable patterns.

Apart from the controversial discussion on the exact size, nature, and layering of the different prosodic units, a further problem is posed by the fact that the status or the range of the prosodic speech units might change in relation to external factors like style, discourse context, or speech tempo (see, e.g., Kleinhenz 1998). Thus, the same (written) string might be realized by several variants of prosodic grouping. Boundaries to indicate prosodic domains (be they prosodic words or phrases) are not necessarily given (see also Section 3.7.2).⁵

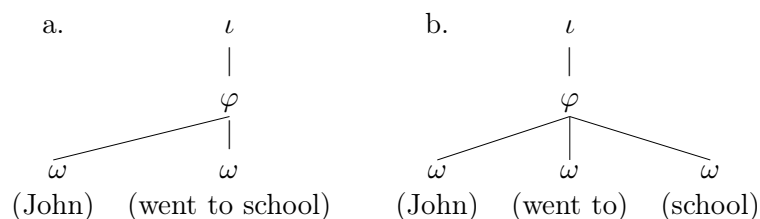


Figure 3.5: Some phrasing possibilities for *John went to school*.

From the discussion above it becomes clear that while the prosodic hierarchy is an essential part of prosodic organisation, it is also a rather widely debated concept with a wide range of variation due to factors from different modules of the grammar (e.g., information structure) as well as external (performance-related) circumstances. The approach pursued in this thesis therefore shifts the focus from the prosodic hierarchy as the scaffold of p-structure to a more general approach, where prosodic constituency is still an important, but not the primary/underlying factor of the p-structure representation.

However, before the approach proposed in this thesis is introduced in more detail, the following section returns to the three representations of p-structure discussed in the previous section: the tree-based representation, the AVM-based representation, and the combination thereof. As will be shown below, these representations are suboptimal when it comes to representing prosodic structure.

⁵These particular considerations are rather part of the *performance* of a specific sentence and are thus not in the focus of grammar theory. Nevertheless, as a concrete implementation is kept in mind here, these facts add to the overall decision. Moreover, note that phenomena like this make it difficult to establish a clear-cut boundary between *performance* and *competence*, as the findings of performance should also be reflected to a certain extent in the abstract representation of language (competence).

3.2.2 The representational problem

The representation of information in a tree or an AVM format or a combination thereof is very familiar to those working within LFG, as it is similar to the well-known representations of c- and f-structure on the syntactic side of the LFG grammar. However, these are not necessarily the best representations for prosodic structure as well.

The tree-based representation as proposed by O'Connor (2004) and Dalrymple and Mycock (2011) relies on hierarchical structures and allows only for a single aspect of the overall prosodic information to be encoded. The most obvious choice for this structure are the elements of the prosodic hierarchy, as they represent other factors of the speech signal on whose basis they are calculated (e.g., boundary tones, pauses, segmental processes, etc.). However, as discussed in the previous section, prosodic phrasing is fickle and is determined by a wide variety of factors. Furthermore, additional relevant p-structure information⁶ has to be stored in a separate AVM structure (Mycock and Lowe 2013) which is again dependent on the correct determination of the inherently varying prosodic hierarchy.

While c-structure is per se hierarchical, the inherent nature of the speech signal is rather linear, and even though prosodic grouping definitely exists and is an important factor, the p-structure representation should also allow for a fine-grained approach to the utterance by enabling the description of various aspects and layers of the speech signal with respect to linear order (i.e., time). This does not exclude prosodic phrasing per se, as there is no denying a certain rhythmic grouping, but the focus is shifted from prosodic constituency to a more balanced representation of all factors relevant in p-structure. Furthermore, the extraction of additional information associated with prosodic structure from a flexible (i.e., shifting) tree-representation might prove to be cumbersome in concrete (computational) implementations.

In contrast to the tree-based representations, the AVM-structure proposed by Butt and King (1998) allows for the representation of a broad spectrum of information in that the attribute value pairs can encode several aspects of a speech signal in one structure. However, typical LFG AVMs, for example f-structure, do not represent information in a linear order. If an AVM-approach is pursued, the AVM must include information on precedence relations; otherwise, an extra ordering instance between string and p-structure is needed.

While AVMs are not suitable to represent linear structures, they are optimal for the indication of nested structures and for constructions where a feature path is required to apply twice, as it is the case with expressions like *She told John to go home*, where *John* is the object of the matrix clause, but the subject of XCOMP. Here, one feature structure (*John*) is shared by two 'positions' in the functional analysis (OBJ and XCOMP SUBJ). However, the only nested constructions that can

⁶For example, which syllable carries nuclear stress, which tone applies within a phonological phrase, etc.

also Zec 2007, Goldsmith 2011, and references therein) and thus forms a reliable basis for a p-structure representation (see also Chapter 2, Section 2.5.2.2).

With the focus on the syllable, the approach presented here is based on the smallest rhythmic unit of the prosodic hierarchy, which has also played an important role in other accounts of rhythm, e.g., in metrical phonology (Lieberman 1975, Liberman and Prince 1977).

Note, however, that the approach can, in principle, be extended to include even smaller units, e.g., morae or segments. However, concerning segments, Nespor and Vogel (1986, 14) state that segments seem to have a “more general freedom [...] to be involved in relations that do not conform to a notion of ‘strict layers’”. In particular they name ambisyllabicity, where a segment belongs to more than one syllable, and the possibility of feature spreading between adjacent segments. Thus, the approach focusses on the syllable as the basic unit, with extensions possible if needed.

Although syllables are considered to be part of the prosodic hierarchy, the approach introduced here is not built upon the general notion of prosodic units and their respective grouping according to hierarchical restrictions, but encodes the events encountered in the speech signal in a linear order. However, since prosodic units are determined by different indicators within the speech signal, these units are accounted for as well, because the required information (on, e.g., breaks or boundary tones) is encoded in the p-structure representation. The syllable as the basic unit also allows for a relative independence from external factors like speech tempo, since the basic unit of the syllable remains the same (in contrast to a time-based approach or an approach based on prosodic grouping).⁷

The following introduction to the p-diagram describes the process from the perspective of *comprehension*, as the p-diagram’s generation is fundamentally different if coming from the perspective of production (see Chapters 4-6). Furthermore, it must be noted that the analysis presented here is by no means an approach to speech recognition. The problem is that p-structure as such cannot be related directly to the speech signal, as there has to be a step in between, some interface between phonetics and phonology. This is, however, a very wide field, and researchers working within this specific area are not agreed upon the exact implementation (for more information, see Kingston 2007, and references therein). A discussion (and inclusion) of the phonetics-phonology interface would go beyond the scope of this thesis, so the p-diagram as proposed in the following must be understood as a coarse representation of the speech signal with syllabic preorganisation.⁸

⁷There might be some *postlexical* syllable reduction or deletion depending on the speech tempo, but either a syllabic consonant remains or the syllable is completely deleted. This does not affect the overall representation. For an example of syllable reduction/deletion, see the discussion on the Swabian data in Chapter 4.

⁸In principle, this representation should not be syllable-based, but should have smaller frames, e.g., segments, which are grouped to syllables as part of a phonological grouping process. However, this additional step from segment to syllable is not relevant for the discussion of the prosody-syntax interface and is thus left for a concrete implementation in the future in favor of a simplified

3.3.1 The making of the p-diagram

The p-diagram in comprehension represents the different aspects of the speech signal in relation to the respective syllables, e.g., fundamental frequency, intensity or duration. This information is extracted from the speech signal and encoded in a *feature vector* that is assigned to the syllable via the following function (from linear algebra):⁹

$$\mathbf{S}(\mathbf{u}(\mathbf{n}), \mathbf{n})$$

Where \mathbf{S} denominates the vector, \mathbf{u} is the vector and \mathbf{n} is the index of the vector from a linear perspective (i.e., in relation to time).

Phrased differently, this function refers to a syllable S , which is assigned a vector (\mathbf{u}) in relation to the linear position (\mathbf{n}) of the syllable in the speech signal. In contrast to a vector as it is used in mathematics, a feature vector is thus not numerical, but consists of discrete dimensions of a specific object. These predetermined dimensions encode a specific information given in the acoustic signal in relation to the respective syllable; a flexible representation, as dimensions can be added or deleted depending on the researcher’s interests. The result is a linearly ordered set of vectors, which include ordered information of different speech signal dimensions in syllable-based frames. These vectors can be generalized as in (18).

$$(18) \text{ S: } \begin{pmatrix} \textit{segments} \\ f_0 \\ \textit{duration} \\ \textit{intensity} \\ \dots \end{pmatrix} (\mathbf{n})$$

The abstract vector in (18) shows four possible dimensions for a feature vector as proposed in this thesis. *Segments* refers to the segments included in the syllable represented in IPA (see Section 3.4.2 for the assumptions made about segments). F_0 refers to the fundamental frequency which describes the frequency of the opening and closing of the vocal folds measured in Hertz and is perceived as the ‘pitch/speech melody’ of a signal. *Duration* refers to the concrete length of the syllable. *Intensity* finally measures the released air pressure in dB, usually perceived as ‘loudness’.

Consider the following acoustic signal of the name *Ravi* (produced for reasons of exemplification): The top part shows an oscillogram, which takes time and amplitude (sound pressure) of the signal into account. The middle part is a spectrogram

syllable-based representation in the present.

⁹In speech recognition, much more fine-grained feature vector extractions are common. For a detailed description see Jurafsky and Martin (2009, 329) who demonstrate the extraction of features from a waveform. The result of the extraction is a 39-dimensional feature vector for every 10 milliseconds in a speech signal. The application of feature vectors is, however, not restricted to speech recognition, but is applied in other areas of (computational) linguistics as well, e.g., machine learning (Liang and Potts 2015, and references therein) or information retrieval (Salton 1971).

in which energy is displayed with reference to time (horizontally) and various frequencies (vertically). The darker a phasis in the spectrogram, the higher the energy density. The lower part allows for an annotation.

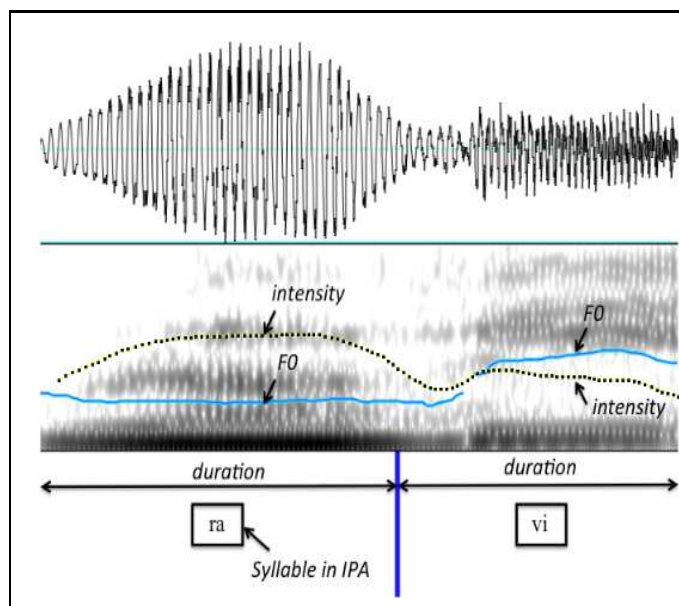


Figure 3.7: Speech signal for *Ravi*.

Applying the abstract feature vector in (18) to the speech signal in Figure 3.7 would yield two feature vectors:

$$(19) \quad S: \begin{pmatrix} [ra] \\ 173 \text{ Hz} \\ 0.2 \text{ s} \\ 78 \text{ dB} \\ \dots \end{pmatrix} (1) \quad S: \begin{pmatrix} [vi] \\ 256 \text{ Hz} \\ 0.16 \text{ s} \\ 68 \text{ dB} \\ \dots \end{pmatrix} (2)$$

Note that the internal order and the magnitude of the vectors extracted from a speech signal need to be identical; i.e., if a certain dimension does not return a value, then a placeholder needs to appear in the vector, otherwise vectors cannot be compared.

Following this extractions process, the set of ordered feature vectors is merged in a representation similar to a diagram, where the syllables encountered in the signal are the basic entities (the ‘x-axis’). On top of these basic entities, the feature vectors are aligned with the respective dimensions forming the ‘y-axis’. Figure 3.8 shows the p-diagram for the feature vectors given in (19).

↑
Feature Vector	↑
	INTENSITY	78	68	...
	DURATION	0.2	0.16	...
	FUND.FREQ.	173	256	...
	SEGMENTS	[ra]	[vi]	...
↑
VECTOR-INDEX	S₁	S₂	...	→

Figure 3.8: The p-diagram for the acoustic signal for [ravi] (Figure 3.7).

Through the alignment of dimensions on the y-axis and vector-indices on the x-axis every value of the p-diagram is accessible. For example, the duration value of the second syllable can be extracted directly from the diagram via the corresponding function $[S_2, \text{DURATION}]$, which would return $\text{DURATION} = 0.16$. For an extraction of all syllable-values of the utterance $[, \text{SEGMENTS}]$ would return a list of all values found at the position of the SEGMENTS of each vector, that is, all x-axis values for this y-axis dimension: $\text{SEGMENTS} = [\text{ra}, \text{vi}]$.

3.3.2 A note on pauses

While the speech signal in Figure 3.7 provided values for every dimension of each feature vector, there are often sequences in a signal without an acoustic content: the pauses. Previous publications on the p-diagram (Bögel 2012, Bögel 2013) assigned individual feature vectors to pauses, where all dimensions were set to zero, except for the duration.¹⁰ If the speech signal in Figure 3.7 above was followed by a pause, the corresponding p-diagram would thus be encoded as in Figure 3.9.

↑
Feature Vector	↑
	DURATION	0.2	0.16	0.2
	FUND.FREQ.	173	256	-
	SEGMENTS	[ra]	[vi]	-
↑
VECTOR-INDEX	S₁	S₂	S₃	→

Figure 3.9: The p-diagram of [ravi]+[pause].

However, separate vectors for syllables and pauses led to awkward annotations at the c-structure–p-structure interface, because a c-structure annotation cannot directly refer to a specific vector in the p-diagram if this vector does not have a value for

¹⁰Bögel (2014) implicitly disposed of these vectors in that intonational phrase boundaries in production were no longer assigned to a separate vector.

SEGMENTS: c-structure does not process ‘empty’ material and does not know about pauses. The rather bumpy solution were references to the (syllable-)vector *before* the pause, followed by a mathematical addition of +1 to the vector index: $S_{2+1}=S_3$.

The current approach disposes of the ‘pause-vector’. Instead, it is assumed that the duration of a pause is calculated earlier and that the corresponding vector is merged with the preceding syllable vector at the phonetics–phonology interface briefly discussed above. In order to separate the syllable duration from the pause duration, another dimension is added: P_DURATION. The resulting p-diagram is shown in Figure 3.10.

Feature Vector	↑	...	↑	↑	...
	P_DURATION		0	0.2			...
	DURATION		0.2	0.16			...
	SEGMENTS		[ra]	[vi]			...
	VECTOR-INDEX		S_1	S_2			... →

Figure 3.10: The p-diagram and the P(AUSE)_DURATION dimension.

So far, the p-diagram is not much more than a coarse-grained representation of the speech signal, with very concrete values to dimensions that are rather part of phonetics than phonology. However, the integration of these values serves as a reminder of a very important fact, namely that any abstract representation/process in phonology is based on the interpretation of concrete speech signal factors – and more importantly, that the mediator at the syntax-prosody interface, the prosodic constituency, is ultimately based on the different values given in the speech signal.

3.3.3 The *interpretation level* of the p-diagram

The rather concrete representation of the speech signal in the previous section is somewhat outside the scope of this dissertation, but it has an important bridging function from the (concrete) speech signal to the (abstract) p-structure representation. In the following, this layer will be called *signal level*. The information included in the signal level is the foundation of the *interpretation level* of the p-diagram, which abstracts away from the concrete information given in the signal level and stores the resulting information for further use at the interfaces to other modules.¹¹

Several parts of the information encoded in the signal level can be represented in a more abstract (and normalized) way.

¹¹A third level of the p-diagram, the *lexical level* which is relevant for production, will be introduced in Chapter 4.

1. The course of the fundamental frequency in Hertz values over time is, for example, often replaced by a calculation of **semitones**, because the higher two Hertz values are, the smaller is the difference between them from a listener perspective. More concretely, this means that the difference between 100 and 150 Hertz is much more significant than the difference between 400 and 450 Hertz. For this reason, semitones and the difference between them are calculated to measure the relative (and not the absolute) difference between two Hertz values.¹² The formula in (20) calculates the semi-tone value for a Hertz-value.¹³

$$(20) \text{f0max(St)} = 12 * \log_2 \left(\frac{\text{f0max(Hz)}}{\text{f0min(Hz)}} \right)$$

The semitone difference between 100 and 150 Hertz will be ca. 7, while the semitone difference between 400 and 450 Hertz will be ca. 2. This difference between semitones gives a better indication of how significant a drop/rise in the fundamental frequency is from a listener perspective. Thus, the p-diagram does not encode the semitone for a particular f_0 value, but the semitone *difference* of the current semitone to the one of the previous element: SEMIT_DIFF. Consequently, a negative value indicates a drop in the fundamental frequency, while a positive value indicates a rise, a representation that allows for a quick overview on significant drops/rises in the speech melody.

2. Closely connected with the representation of semitone differences are the **ToBI** conventions for labelling High and Low tones and break indices in intonation, thus modelling the pitch contour of a speech signal (Silverman et al. 1992). The English ToBI inventory includes two monotonal (H^* , L^*) and three bitonal pitch accents ($L+H^*$, L^*+H , $H+!H^*$), as well as boundary tones (Beckman and Hirschberg 1994). The German ToBI inventory (GToBI) includes the same inventory, but features one more pitch accent: $H+L^*$ (Grice and Baumann 2002).

3. **Break indices** are also part of the ToBI annotation and indicate the value of the perceived degree of disjuncture between two words (Beckman et al. 2005). As such, they constitute an abstract representation of P_DURATION as discussed above. Break indices range from 0 (clitic boundary) to 4 (intonational phrase), thus grouping segments of speech hierarchically. For the current version of the p-diagram, only break indices 3 (phonological phrase) and 4 are taken into account.

¹²To give an assessment of semitones: 12 semitones constitute an octave.

¹³There are several possibilities to calculate semitones: First, the maximum value of the current syllable relative to the minimum value of the fundamental frequency of the utterance. This will apply for all subjects, independently of their pitch range. Second, the semitones can be calculated relative to, e.g., a 100 Hz, in case the minimum value cannot be determined. A 100 Hz will be sufficient for female participants; for male participants, the value should be set to 50 Hz or lower. However, while the absolute semitone values differ with the two formula options, the p-diagram only encodes the difference between two semitone values. The value of this difference is the same with both options.

4. **Phrasing** finally indicates prosodic constituency and is determined by various factors. In comprehension, PHRASING indicates larger (and thus more easily identifiable) constituents, i.e., the phonological phrase and the intonational phrase. These are calculated on the basis of different parameters, e.g., the break indices, the boundary tones, abrupt resetting of the intonation, or the duration of the previous syllable.

In order to give a concrete example, consider the following speech signal encoding the coordination *Ravi and Amra or Karla* as a complete utterance.

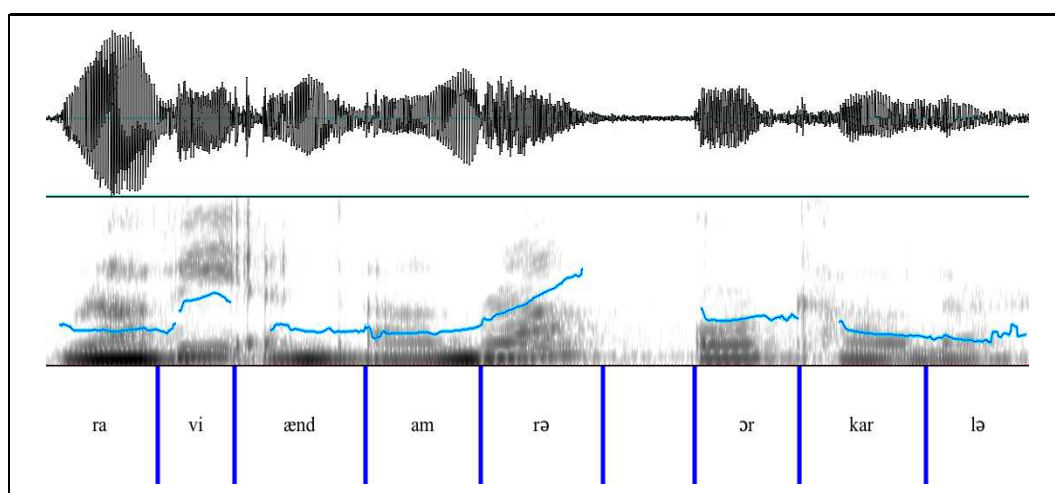


Figure 3.11: Speech signal for *Ravi and Amra or Karla*.

The acoustic signal represented in Figure 3.11 is encoded in the following p-diagram:

↑ ...	↑	↑ Interpretation
PHRASING	-	-	-	-	φ	-	-	ι	↓
SEMIT_DIFF	..	6.8	-6.8	-0.4	7.5	-4.6	-3.2	-2	
TOBI	L*	+H		L*	+H H-		L*	L-L%	
BREAK_IND	-	-	-	-	3	-	-	4	
...	Signal
DURATION	0.25	0.16	0.28	0.24	0.26	0.23	0.27	0.23	↓
P_DURATION	0	0	0	0	0.2	0	0	0.4	
FUND._FREQ.	173	256	174	170	263	201	167	149	
SEGMENTS	[ra]	[vi]	[ænd]	[am]	[rə]	[ər]	[kar]	[lə]	
VECTORINDEX	S₁	S₂	S₃	S₄	S₅	S₆	S₇	S₈	...

Figure 3.12: P-diagram for the speech signal in Figure 3.11.

In addition to the *signal* information, the p-diagram in Figure 3.12 also encodes the *interpretation level* on the basis of the speech signal data as it was discussed above. At positions S₅ and S₈, pauses and boundary tones lead to an indication of prosodic phrasing: A phonological phrase boundary after S₅, and an intonational phrase boundary after S₈.

With the prosodic units encoded under PHRASING, the p-diagram allows for a representation of the signal in terms of the prosodic hierarchy, even though it is not, per se, hierarchically constructed. However, in contrast to the p-structure representations discussed in Section 3.2, the p-diagram allows for a stable representation of the data independently of prosodic constituency while at the same time providing a linear and compact representation of p-structure. Furthermore, to a certain extent, it is also a ‘neutral’ representation: The calculation of the *interpretation level* is dependent on the individual researcher (and the associated theory).

This section established the p-diagram during comprehension. However, the question remains as to how this information is relevant for the central topic of this thesis: the prosody–syntax interface in LFG. The first issue that needs to be discussed in this context is the *transfer of vocabulary*: How are the segments of the syllables in the p-diagram associated with the corresponding morphosyntactic units? The answer to this question lies with the multi-dimensional lexicon as it is introduced in the following section.

3.4 The multi-dimensional lexicon

The lexicon model proposed in this thesis roughly follows the assumptions of Levelt et al. (1999) in that the lexical entry has several aspects that belong to different modules of grammar. Within LFG, Dalrymple and Mycock (2011) have proposed an extension to the standard morphosyntactic entry (the s(syntactic)-form), which they call the p(honological)-form and which includes information on syllable structure, syllable weight and an IPA transcription as represented in Figure 3.13 for the lexical entry of the word ‘university’.

<i>s-form</i>	university
<i>f-description</i>	N (↑PRED)=‘university’
-----	-----
<i>p-form</i>	/ju _w .ni _w .və:s.si _w .ti _w /

Figure 3.13: Lexical entry for ‘university’ (Dalrymple and Mycock 2011).

The purpose of such an extension is self-evident: it allows for an association of a string of segments with a corresponding syntactic form. Thus, the following sections will take this approach towards a lexical p-form and extend it with material proposed by Levelt et al. (1999).

3.4.1 The lexicon according to Levelt et al. (1999)

In their model, Levelt et al. (1999) assume three different strata to be relevant for a lexical item. First, the *conceptual* stratum describes the semantic value of an item and the activation of related lexical items (hyperonyms etc.). For example, if the conceptual stratum ESCORT is activated, this activation will also prime related words like SAFEGUARD and ACCOMPANY. Such a (semantically) primed activation can be measured by lexical decision tasks, where a participant has to decide if a word is a lexical word. If a participant is (e.g., orally) presented with the word *apple*, he/she will decide significantly faster for *fruit* to be a lexical word in comparison to *ship*, because *ship* is lexically unrelated to *apple* and has thus not been primed beforehand (see, e.g., Friederici et al. (1999) and references therein).

Connected to the conceptual stratum is the *lemma* stratum, which includes the lexical category and the morphosyntactic information associated with a lexical item. For the lemma *escort*, the lexicon will store the lexical category information that *escort* is transitive. Furthermore, information on how, e.g., tense and aspect are realised are also stored along with the lemma.

Depending on the information stored in the lemma, the phonological form of the lexical item will be realised in the *form* stratum. This last stratum includes several aspects concerning the item's articulatory realisation. One is the word's morphological makeup, i.e., the morphological realisation of abstract lemma information. For example, if the lemma *escort* is associated with the information TENSE=pres and ASPECT=prog, two morphemes are involved: <escort> and <ing>. The resulting phonological realisation is /ɛskɔrtɪŋ/.

Furthermore, the *form* stratum spells out the metrical and segmental properties of these morphemes. Figure 3.14 shows the (slightly modified and reduced) form stratum as proposed by Levelt et al. (1999, p4, fig. 2) for the word 'escorting'.

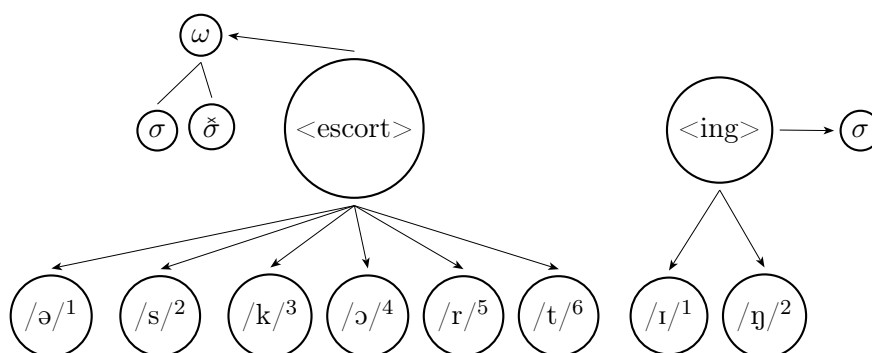


Figure 3.14: Form stratum as proposed by Levelt et al. (1999)

Each morpheme receives a metrical frame which includes stress, the number of syllables, and information about the prosodic status of an element (e.g., is it a prosodic

word or merely an unstressed syllable). For the morpheme <escort>, this frame indicates a disyllabic, iambic prosodic word. The morpheme <ing> on the other hand only consists of one unstressed syllable and is dependent on a prosodic host, i.e., it needs to attach to another element to form a prosodic word.

Each morpheme is associated with an ordered set of linearly ordered segments. These segments are not grouped into syllables, because the distribution may change in accordance to the phonological environment of the lexical item. In the above example, <escort> by itself would form two syllables, with the /t/ phoneme being in the coda of the second syllable: [ə.scɔrt]. However, if the present progressive morpheme <ing> is attached to the stem, the /t/ will inhabit the onset position of the next syllable because of onset maximisation: [ə.scɔr.tɪŋ]. It is thus important to note, that while the lexicon according to Levelt et al. (1999) stores metrical frames and phonemic segments, it does not group them together to spell out the syllables associated to a lexical item.

3.4.2 The lexical model proposed in this thesis

The lexicon as proposed by Levelt et al. (1999) does not only present the ‘surface’ form, but discusses how this form is derived during the process from MEANING (‘concept’) to FORM at the word level, a discussion which goes deeply into the creation of words and is beyond the scope of this thesis.¹⁴ Instead, this thesis focusses on the information that is the *output* of each stratum. For this proposal, parts of Levelt et al.’s threefold lexicon model are adapted (thus extending the lexical model proposed by Dalrymple and Mycock 2011) as it provides every aspect necessary for a thorough analysis at the interface between prosody and syntax.

1. The **concept**: The concept describes the semantic concept of a lexical item; i.e., the *idea* of an entity in the world before a lemma (a morphosyntactic item) and a form (a phonological representation) are added to express this idea. However, even though the concept is part of the lexical representation, its exact nature is not discussed here, as the focus of this thesis is on the prosody-syntax interface and thus on the information provided by the *lemma* and the *form*. A discussion of the information related to the concept is left to future research on, e.g., the prosody-semantics/i-structure-interface.
2. The **lemma**: The lemma is the morphosyntactic representation of the concept. Following Butt and Kaplan (2002), its realisation is a fully formed unit of the s-string/a terminal node of the c-structure. It carries morpho-syntactic information on, e.g., word category, tense or subcategorization frames, which is crucial for a correct syntactic interpretation. Following Dalrymple and Mycock (2011), the lemma is referred to as *s(yntactic)-form* in this thesis.

¹⁴An underlying model similar to the ones proposed in the tradition of lexical phonology and morphology (among others, Kiparsky 1982b) is assumed, but not defended here. See, however, Chapter 4 for a discussion on *postlexical* phonology in the same tradition.

3. The **form**: The form describes the phonological representation of the concept and its morphosyntactic representation. In this thesis, it is referred to as *p(honological)-form* (cf. Dalrymple and Mycock 2011). The p-form entry encodes information about the *metrical form*, i.e., syllable structure, stress, and prosodic phrasing. Furthermore, the single *segments* are given.¹⁵ Together, these attributes contribute to the p-form as it is represented in isolation and before the application of postlexical processes.

S-form and p-form are two different aspects of the same concept; they represent two sides of one coin, and while they may look very different and encode very different information, they still represent the same core – their concept. In (21), the simplified morphosyntactic entry (s-form) of the name *Amra* is given.

(21)	Amra	N	(↑ PRED)	= ‘Amra’
			(↑ GEND)	= fem
			(↑ NUM)	= sg
			(↑ PERS)	= 3
			(↑ NTYPE)	= proper_name

In short, this entry states that there is a lexical entry *Amra*, which is a noun (N) and which has predicate status, feminine gender, singular number, third person, and the noun-type is a proper name. *Amra* is a morphologically ‘simple’ word. Note, however, that morphologically ‘complex’ words, e.g., *walks*, do not consist of two lexical entries *walk* and *-s* in LFG. Instead, the difference between the present tense form of *walk* and *walks* is indicated by a difference in the features PERS and NUM: *walks* is the third person singular form while *walk* represents all other combinations – a process of word formation contributed to morphology. The lexicon’s *output*, however, as it is accessed by the syntax, is assumed to consist of *morphologically complete* words (as discussed in, e.g., Butt and Kaplan 2002, and essential to the lexicalist hypothesis as applied in LFG).

While lexical entries of s-forms (as in (21)) have been widely used in theoretical and computational LFG grammars, the p-form as it is proposed in this thesis requires further explanation. If dealing with isolated words, the syllables are clearly set in the mind. However the syllables’ clear boundaries may be lost during runtime as resyllabification might take place, where the segment originally placed in the coda of the previous syllable is drawn to the onset of the next syllable/word. This is one of the reasons why Levelt et al. (1999) propose that the lexicon stores the segments of a morpheme in an ordered set, but does not group these segments

¹⁵Note that one lexical entry can include many p-forms, as they may differ according to dialect, speech register and other external reasons. This thesis follows Lahiri and Reetz (2002, 2010) who assume an abstract and underspecified representation of phonological features, which allows for a flexible treatment of different phonetic representations of the same concept. However, in order to simplify the p-form representation in the lexicon, the IPA ‘standard’ pronunciation is used.

into syllables. Instead, a *metrical frame* is stored along with the segments, which contains information on the number of syllables and the distribution of stress, but allows for flexibility in the placement of the segments depending on the phonological environment. Furthermore, the prosodic status is given as well. For the s-form *Amra* ((21)), the p-form representation would thus look like the following:¹⁶

$$(22) \quad [{}^1\text{am.r}\partial] \quad \begin{array}{l} \text{SEGMENTS} \quad \quad \quad /a \ m \ r \ \partial/ \\ \text{METRICAL FRAME} \quad \quad ({}^1\sigma\sigma)_\omega \end{array}$$

In principle, morphologically complex words could be represented by several p-form entries that interact with each other according to specific phonological rules. Gussenhoven (1991, among others), for example, discusses the derivation of the noun or adjective *Japanese* by means of suffixation of the stressed derivational affix *-ese* to the noun *Japan*. The corresponding p-forms are given in (23).

$$(23) \quad \begin{array}{ll} \text{a.} & [d\text{ʒ}\partial.\text{p}\partial\text{en}] \quad \text{METRICAL FRAME} \quad (\sigma^1\sigma) \\ \text{b.} & [{}^1\text{i:z}] \quad \text{METRICAL FRAME} \quad ({}^1\sigma) \end{array}$$

The interesting fact is that, taken together, the stress distribution changes, as two stressed syllables are placed next to each other. The English rhythm rule (Gussenhoven 1991, Shattuck-Hufnagel et al. 1994) will then in return cause the stress on the second syllable of *Japan* to move to the first syllable, which results in a new metrical frame.

$$(24) \quad [d\text{ʒ}\partial.\text{p}\partial.\text{ni:z}] \quad \text{METRICAL FRAME} \quad (\sigma^1\sigma)+({}^1\sigma) \Rightarrow ({}^1\sigma\sigma^1\sigma)$$

It is important to stress the fact that while it is assumed that the lexicon is created dynamically through morpho-phonological processes, this thesis (and the interface processes discussed therein) *only* accesses the **surface representations** of the lexicon. For the two lexical items *Japanese* and *Amra*, this results in the following representation of the lexicon, as it is used throughout this thesis.

¹⁶Note also that while the lexical entry separates metrical (syllable) structure and segments, the p-form itself is represented by a fully syllabified structure, although postlexical syllabification is assumed. The reason for this is twofold: a) the p-form represents the lexical item as it is pronounced in isolation, and b) the transfer of lexical syllables simplifies the representation from the perspective of production. This will be discussed in more detail in Chapter 4.

concept	s-form	p-form
JAPANESE	S-FORM <Japanese>	P-FORM [dʒæ.pə.'ni:z]
	N (↑ PRED) = 'Japanese' (↑ NUM) = pl	SEGMENTS /d ʒ æ p ə n i: z/ METRICAL FRAME (σσ'σ) _ω
	...	
	A (↑ PRED) = 'Japanese' ...	
AMRA	S-FORM <Amra>	P-FORM ['am.rə]
	N (↑ PRED) = 'Amra' (↑ NUM) = sg	SEGMENTS /a m r ə/ METRICAL FRAME ('σσ) _ω
	...	
	...	

Table 3.1: Lexical entries for *Japanese* and *Amra*.

By adding an additional phonological layer to the lexical entry the lexicon can be used as a lookup device which aligns strings of the speech signal with the corresponding morphosyntactic items in a dynamic process (as, e.g., described by McQueen (2005) and references therein). This alignment is an essential part of the prosody–syntax interface; especially in cases where a lexical disambiguation helps to differentiate between homographic s-form representations. In English, for example, lexical stress can differentiate between categories (e.g., *'permit* (noun) vs. *per'mit* (verb)); thus, depending on the position of lexical stress, a specific lexical entry is chosen and processed by syntax as either a noun or a verb with its respective c- and f-structure representations.

The same is true for German, where the representation of lexical stress can be essential in order to distinguish between two syntactic/conceptual representations. The following sentence is ambiguous in its isolated written form in that the reader cannot discern between the (otherwise separable) particle verb ('to cross over') and the inseparable verb ('to translate').¹⁷

- (25) **S-string:** Lass uns übersetzen.
'Let us translate / cross over.'

If, however, a speech signal is given, then the phonological information is essential: While the s-forms of the two concepts are homographic in verb-final sentences, the respective p-forms show a difference (Table 3.2). If lexical stress is found on the first syllable, the associated concept would be 'to cross over'. If, on the other hand, the lexical stress is on the third syllable, the concept would be 'to translate'.

¹⁷See also Chapter 2, Section 2.5.2.2 for a brief introduction into the treatment of particle verbs in this thesis.

CONCEPT	S-FORM	P-FORM
<i>translate</i>	übersetzen (V)	[y:.bɛ.'zɛts.ŋ]
<i>cross over</i>	übersetzen (Particle-V)	[y:.bɛ.zɛts.ŋ]

Table 3.2: Abbreviated lexical entry for *übersetzen* ('to translate'/'to cross over').

If this information on word accent is available 'outside' of the lexicon (e.g., in form of a speech signal representation), a disambiguation of meaning can take place during the alignment process and multiple syntactic and semantic interpretations can be avoided. An exclusively syntactic analysis of example (25), on the other hand, cannot differentiate between the two verbs.¹⁸

3.4.3 The *transfer of vocabulary*

Having established the lexical model and how it aligns phonological and syntactic forms in the previous section, the output of the p-diagram in Figure 3.12, referring to the speech signal of the coordination *Ravi and Amra or Karla*, can be aligned first with the corresponding p-forms and subsequently with the associated s-forms.

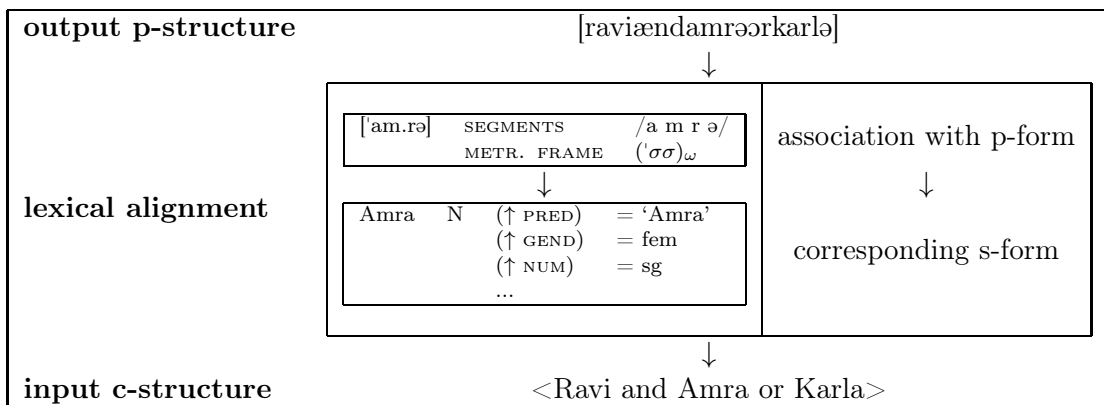
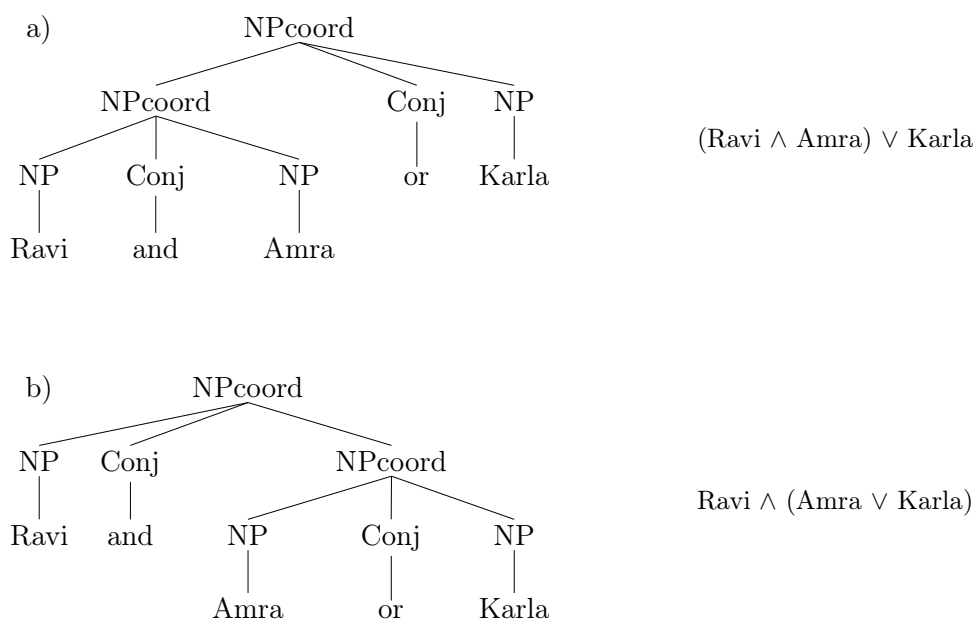


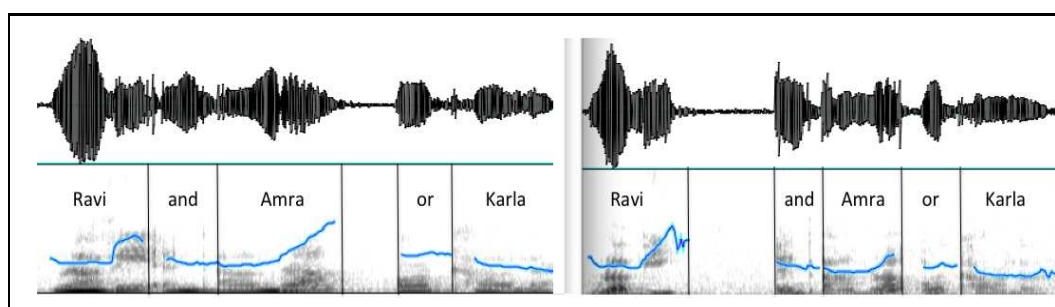
Table 3.3: Lexical alignment: the *transfer of vocabulary*.

This *transfer of vocabulary*, which aligns p- and s-forms with each other and thus enables a communication between the modules on a lower level, is an essential part of the interface communication. It is reversible in that the minimal units of c-structure (the s-forms) can be associated with the corresponding p-forms during production (see the following chapters). The result of this *transfer of vocabulary* is the input to c-structure: *Ravi and Amra or Karla*. Parsing this coordination will return the c-structure representations in Figure 3.15.

¹⁸See also Forst et al. (2010) who analyse particle verbs in LFG. This was briefly discussed in Chapter 2, Section 2.5.2.2. Note that no particular theory concerning the lexical/syntactic treatment is assumed here. Particle verbs were chosen because of their orthographic ambiguity with 'full' verbs.

Figure 3.15: C-structure representations for *Ravi and Amra or Karla*.

Wagner (2005) discusses coordination structures like the ones given in Figure 3.15 at length from a syntactic, prosodic and semantic perspective. In these coordination groupings, differences in syntactic grouping are reflected by semantic grouping. That is, the syntactic tree given in (a) groups *Amra* and *Ravi* in opposition to *Karla* while the tree in (b) semantically groups *Ravi* with either *Amra* or *Karla*. As Wagner shows, this relationship is also reflected in prosody in that there is a stronger boundary after the grouping in semantic constructions like (a), and a stronger boundary before the grouping in constructions as in (b).¹⁹ This is also true for the coordination data of Figure 3.15.

Figure 3.16: Speech signal options for *Ravi and Amra or Karla*.

¹⁹In order to verify Wagner's claim, three subjects were asked to group names according to previously established semantic contexts in a small production experiment of a total of 15 items. Every single produced pattern was in accordance with the predictions made by Wagner (2005).

The left speech signal in Figure 3.16 corresponds to the syntactic tree in figure 3.15 (a) and has the semantic form $(Ravi \wedge Amra) \vee Karla$. The speech signal on the right corresponds to the tree in (b) and has the semantic form $Ravi \wedge (Amra \vee Karla)$. Both speech signals reflect the syntactic and semantic phrasing in prosody by means of intonation and pause, or, in other terms, a phonological phrase boundary.

The question is how this difference in prosodic constituency is communicated to c-structure. As the phonological phrase boundary is assigned postlexically to the coordination construction, the assignment of higher levels of prosodic constituency cannot be part of the lexicon and, as a consequence, cannot be part of the (lexical) *transfer of vocabulary*, but must be communicated differently. This second form of interface communication will be discussed in the following section.

3.5 The *transfer of structure*

The underlying architectural assumption in this thesis is that prosody and syntax each have an independent set of module-internal rules and principles as well as a relation with each other that allows for a communication of relevant information. One of the central links between these modules has been established in the previous section: the multi-dimensional lexicon. However, there must be a second level of exchange with information on structural representations, as the example on coordination grouping in the previous section showed.

As has been briefly mentioned in Section 3.1, prosody only has a limited influence on syntactic phrasing. Evidence for this assumption comes (among others) from studies on *garden-path* effects. Garden-path sentences are grammatically correct sentences whose beginning will lead the reader to an incorrect interpretation as demonstrated in (26).

(26) Without her contributions would be inadequate. (Bader 1996, 250)

In sentences like (26), the reader is likely to *not* include an intonational phrase boundary after *her*. Instead, the reader will group *contributions* syntactically and, as a result, also prosodically together with *her* ([her contributions]), before realising that *her* is not a determiner, but a pronoun.²⁰ Following this, the syntactic and, again as a result, the prosodic structure have to be reanalysed (her|[contributions]). It is thus *always* the syntactic analysis that has the final say. It is only when syntax is undetermined between two or more possible analyses that prosody can influence a particular decision. This is also supported by the fact that non-isomorphism between prosodic and syntactic structures, as it frequently occurs especially with function

²⁰Such drastic garden-path effects are, however, reduced to reading experience and to constructions that involve a rearrangement of both, the prosodic and the syntactic component. As Bader (1996, 302) notes, the repair mechanism has to apply to two modules and is thus perceived as being more noticeable than repair strategies that are confined only to the syntactic component. See Bader (1996) for more information on these effects.

words (see, e.g., the *ezafe* construction in Figure 3.1), has no influence on syntactic structuring as well.

For this reason, this thesis does not assume a constant projection of prosodic structure to syntactic structure, as the permanent alignment of syntactic with prosodic constituents is a) not justified from the perspective of syntactic phrasing, as it is the ‘more powerful’ of the two, and b) not justified by experimental studies: Neither do all speakers produce prosodic cues, nor do listeners always understand these cues (Allbritton et al. 1996, Price et al. 1991, see also Section 3.7). Instead, prosodic structure is assumed to be built up independently and it is only in cases of syntactic indetermination that information on prosodic constituency can be requested. This process of ‘requesting’ and the resulting architecture is discussed in the following.

This thesis assumes the (s-)string to be positioned between c- and p-structure (Figure 3.17).

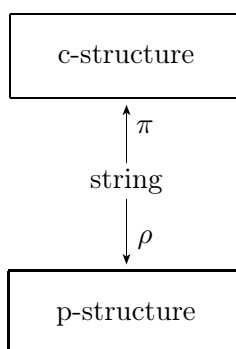


Figure 3.17: The prosody–syntax interface in LFG (first version).

While it is uncontroversial that the correspondence function π relates string and c-structure, the correspondence with p-structure has been introduced in several different positions (see Chapter 2, Section 2.6). This thesis assumes the projection ρ to apply between the string and p-structure as shown in Figure 3.17, which is similar to the proposal of Dalrymple and Mycock (2011) and Mycock and Lowe (2013) (who denominate the relation with β).

Part of ρ 's projection information is the *transfer of vocabulary*, i.e., the projection ρ ‘includes’ a look-up and alignment of lexical information as it was described in the previous section. Note that while lexical access is assumed to be available between p-structure and c-structure, this thesis does explicitly *not* exclude other possible architectural ‘positions’ for lexical access. If, however, the lexicon is accessible everywhere or if it is accessible only in between the modules or if the position proposed above is indeed the only ‘access position’ is left for further research.

As has been stated above, c-structure also has to have direct access to information on prosodic constituency stored in p-structure in order to resolve syntactic ambiguities. This correspondence between c-structure and p-structure can be defined as

the composition (Kaplan 1995, Asudeh 2006) of the inverse relation between string and c-structure (π^{-1}) and the relation of the string with p-structure: $\rho(\pi^{-1}(f))$. In order to simplify the reference to this projection within the phrase structure rules, the following abbreviation is used:²¹

$$\natural(f) \equiv \rho(\pi^{-1}(f))$$

Via this path description, the respective c-structure rule annotation can ‘look up’ or ‘project’ information from and to the related section of the p-diagram. Thus, Figure 3.17 needs to be extended:

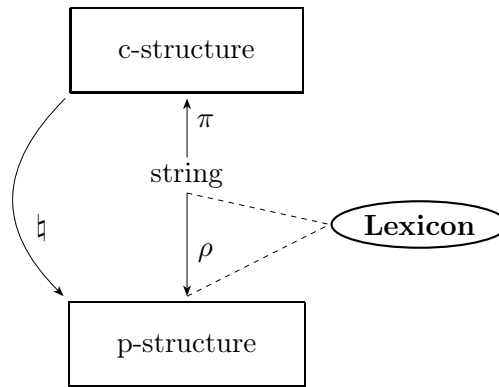


Figure 3.18: The prosody–syntax interface in LFG (second version).

Two interface processes have now been established: the *transfer of vocabulary*, which aligns strings given in the p-diagram with their corresponding p-forms and their associated s-forms, and the *transfer of structure* which allows for the disambiguation of syntactic structures on the basis of prosodic constituency. In the following section, these transfer processes will be applied to coordination grouping.

3.6 Coordination grouping: an analysis within LFG

This section focusses on the analysis of the coordination as given in Figure 3.15 (b) with the semantic representation $Ravi \wedge (Amra \vee Karla)$. The corresponding speech signal includes several indicators of a phonological phrase boundary after the first constituent (*Ravi*) (cf. Figure 3.16). This phrase boundary is calculated on the basis of several factors given in the speech signal, e.g., the pause or the pattern of the fundamental frequency. Figure 3.19 shows the relevant part in form of a p-diagram.

²¹The musical ‡ symbol was chosen, because it represents the relation between c- and p-structure visually with two horizontal lines and two vertical lines in the middle as two parallel, but interacting modules of grammar.

↑ PHRASING	-	φ	-	-	-	↑ interpretation
TOBI	L*	+H H-	-	L*	+H	↓
BREAK_IND	-	3	-	-	-	
DURATION	0.22	0.26	0.26	0.29	0.17	signal
P_DURATION	0	0.55	0	0	0	↓
FUND. FREQ.	181	268	183	164	212	
SEGMENTS	[ra]	[vi]	[ænd]	[am]	[rə]	
VECTORINDEX	S₁	S₂	S₃	S₄	S₅	...

Figure 3.19: Fragment of the p-diagram for *Ravi* \wedge (*Amra* \vee *Karla*).

Based on the data given in the speech signal, a phonological phrase boundary can be assumed after the first syntactic constituent. One dominant factor is the break after the string *Ravi* ($[S_2, P_DURATION/BREAK_IND]$). Further indication comes from the long, even though unstressed, second syllable ($[S_2, DURATION]$) and the boundary accent on this syllable ($[S_2, FUND.FREQ/TOBI]$). Taken together, these single factors are very strong indicators of a phonological phrase boundary ($[S_2, PHRASING]$).

However, in order for this information on prosodic constituency to be instrumental for the disambiguation of syntactic structures, syntax must be given access to this interface information. This can be accomplished via a c-structure annotation with the relation \natural ($\equiv \rho(\pi^{-1}(f))$) that refers to the information accessible at the prosody–syntax interface. The (simplified)²² NP-coordination rule given in (27) represents the interpretation $[Ravi \wedge (Amra \vee Karla)]$.

$$(27) \text{ NPcoord} \rightarrow \begin{array}{c} \text{NP} \\ (\uparrow_{\natural} S_{max} \text{ PHRASING}) =_c \varphi \\ \dots \end{array} \quad \text{Conj} \quad \text{NPcoord}$$

The NP node is annotated with a constraint that refers to p-structure via the correspondence relation path (\uparrow_{\natural}). This is followed by a reference to the relevant vector index, in this case the variable S_{max} . S_{max} corresponds to the vector with the *maximum* index in the scope of the annotated node NP which includes all terminal nodes that are daughters of NP. S_{max} is thus a placeholder for the *last syllable* in a fragment of the (syntactic) string. A similar placeholder is S_{min} which represents the *first* syllable in a string fragment determined by syntax.

²²The syntactic NP coordination rule as such is recursive and furthermore, the choice of syntactic/semantic interpretation also relies on the choice of the conjunction (*or* in combination with *and* will lead to syntactic ambiguities, in contrast to constructions that have two identical conjunctions). The structures presented are thus reduced to a minimum as they are mainly used for demonstration purposes. For a thorough study of conjunction grouping, the reader is referred to Wagner (2005).

- (28) a. S_{max} : A placeholder for the *last* syllable in the phonological representation corresponding to the set of terminal nodes within the scope of the annotated syntactic node.
- b. S_{min} : A placeholder for the *first* syllable in the phonological representation corresponding to the set of terminal nodes within the scope of the annotated syntactic node.

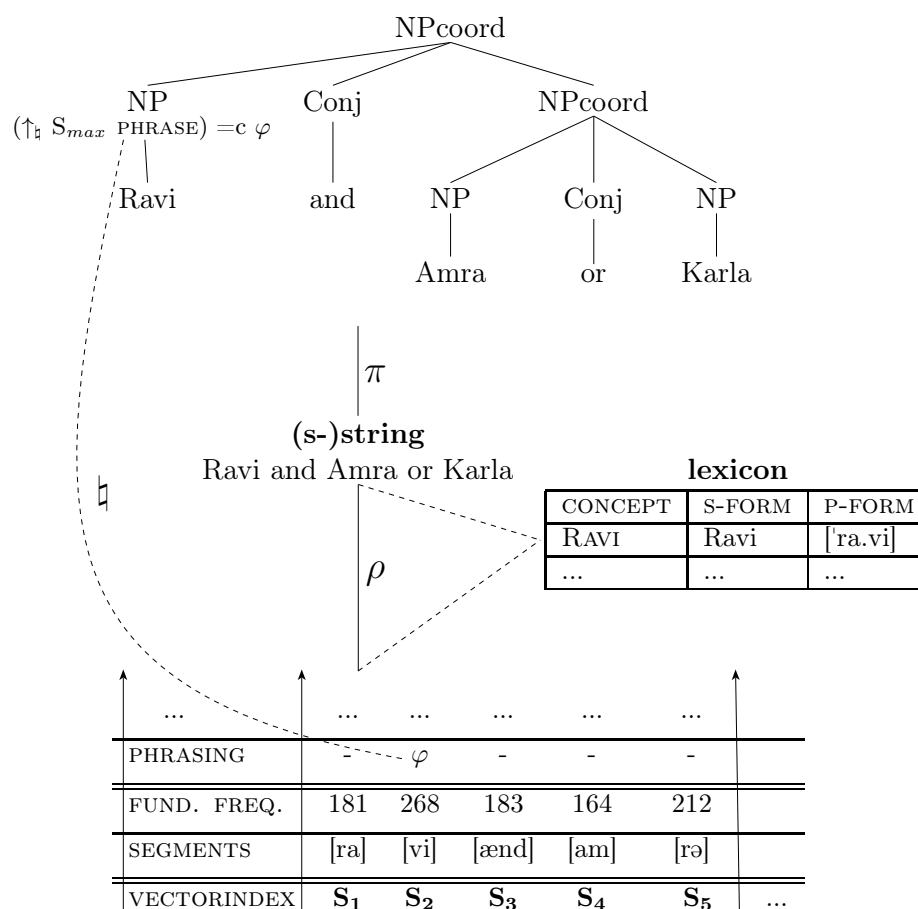
In the coordination example above, NP would only include one terminal node, *Ravi*. The maximum syllable would then be *vi*, which corresponds to the second vector S_2 in the related p-diagram in Figure 3.19. After the relevant vector has been established, the annotation in (27) checks for the attribute PHRASING and states that this attribute *must* indicate a phonological phrase boundary (=c φ). As this is the case, the c-structure rule applies and the correct syntactic tree is returned.

Note that this reference to S_{max} and S_{min} is, in principle, an implementation of the *end-based* approach to the syntax–prosody interface (Selkirk 1986, Chen 1987, see also Chapter 2, Section 2.5.2.4), as it only transfers information on the edge of the syntactic constituent. However, the annotation at the c-structure node can be easily adjusted to represent the *match* approach as proposed in Selkirk (2011): In that case, S_{min} and S_{max} both have to be part of the annotation (see Chapter 6 for an example).²³

In the case presented here, this distinction is not important, as the data is discussed from the viewpoint of *comprehension*: No default mapping from prosodic structure to ‘create’ syntactic structure is assumed (see also Section 3.1 and 3.5, as well as Section 5.4.5 in Chapter 5). Instead, the potentially ambiguous c-structure node ‘checks’ for a structure indication given by p-structure.

Figure 3.20 shows the complete architecture of the prosody–syntax interface for the semantic representation $Ravi \wedge (Amra \vee Karla)$.

²³In fact, in the case of production, the coordination construction would be a problematic case for both approaches to the interface: the *end-based* approach would place a boundary after each noun, not reflecting any recursion (which is wrong, see Wagner (2005)) while *match* would create a highly recursive structure. Furthermore, both mappings would incorrectly phrase *and* with *Ravi*. However, as production is *not* in the focus here, this is left for further research.

Figure 3.20: $Ravi \wedge (Amra \vee Karla)$: at the prosody–syntax interface.

While the prosodic cues given in the acoustic signal of coordination grouping are relatively reliable in that a phonological phrase boundary can be determined, such a clear indication is often not given. As the c-structure annotation given in (27) is a so-called ‘hard’ constraint in that it only allows for an application or non-application of the associated c-structure rule, this irregular occurrence of prosodic cues is problematic as ambiguous structures are no longer allowed. As the following case study on German dative-genitive ambiguities shows as well, it is thus necessary to always encode references to p-structure with soft constraints.

3.7 Case study: German case ambiguities

Discourse context usually calls for a single reading and speakers and listeners can use prosodic information to clarify the meaning of syntactically ambiguous sentences. This section reports on a production experiment with fully ambiguous constructions

that alternate between a genitive and a dative reading as exemplified in (29).

- (29) Überraschend antwortete [der Diener] [der Gräfin]
 Surprisingly answered the.MASC.NOM servant the.FEM.GEN/DAT duchess
 ‘Surprisingly, the duchess’s servant answered // the servant answered the duchess.’

The aim of this experiment was to identify relevant prosodic cues that mark a specific interpretation. In theoretical analyses of speech phenomena, these concrete cues are usually replaced by the abstract notion of a prosodic domain (boundary). However, as this thesis is also written from the perspective of a possible computational implementation in the future, the focus here is on the single concrete cues that indicate such an abstract domain, in this case a phonological phrase. Thus, in the following, the significant cues are considered as well as the ones that are insignificant from a statistical perspective, but are still valuable from the viewpoint of machine processing.²⁴ The underlying question from a theoretical perspective is the question as to how much a ‘real act of performance’ contributes to the determination of rules and constraints that form the core of the grammar.

German case is mainly encoded by the determiner system. In Table 3.4, the determiners for the three genders and four cases of German are listed (in the singular).

	masculine	feminine	neuter
nominative	der	die	das
genitive	des	<i>der</i>	des
dative	dem	<i>der</i>	dem
accusative	den	die	das

Table 3.4: The German determiner system.

Note that there is a syncretism between the feminine form of the dative and the genitive²⁵, which leads to the type of ambiguity illustrated in (29) and below in (30), where the ambiguity in the subordinate clause is based on the ambiguous feminine article of the second DP *der Freundin* ‘of/to the friend’. This DP could either be dative or genitive.

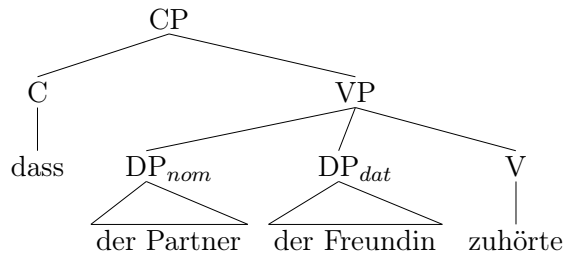
²⁴The aim of a computational implementation would be to cover as many speakers as possible and to reduce ambiguities to a minimum to avoid the application of additional processing power.

²⁵The syncretism is also true for the masculine nominative form. However, this fact has no impact on the ambiguity of the construction analysed in this section.

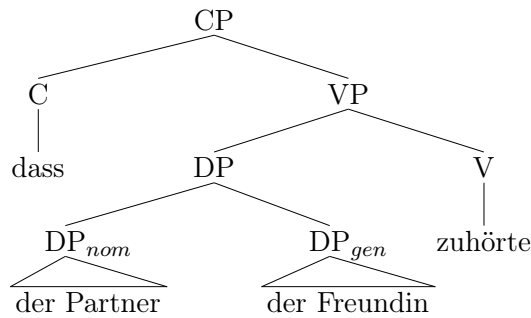
- (30) Alle waren überrascht dass
 Everyone was surprised that
- [der Partner]_{DP1} [der Freundin]_{DP2} zuhörte
 the.ART.MASC.NOM partner the.ART.FEM.GEN/DAT friend listened
- ‘Everyone was surprised that [the friend’s partner listened //
 the partner listened to the friend].’

The sentence in (30) is fully ambiguous, because the final verb is intransitive, but optionally allows for a dative object as well. This full-sentence ambiguity results in two possible c-structures: In (31a), the two DPs are independent daughters of the VP, while in (31b), the two DPs form a possessive construction.²⁶

- (31) a. **Dative:** The partner listened to the friend



- b. **Genitive:** The friend’s partner listened



While the purely syntactic analysis of ambiguous structures leads to multiple representations, a distinction is possible on the basis of prosodic cues (Lehiste et al. 1976, Price et al. 1991, among others).

For their experiment on the importance of different perceptual prosodic cues in German, Gollrad et al. (2010) used sentences with a genitive/dative ambiguity as well. However, their study did not involve completely ambiguous structures. Instead,

²⁶The c-structure analysis is based on the German ParGram grammar (Dipper 2003).

the sentences consisted of three determiner phrases whose relation was disambiguated by the final verb. These sentences were first used in a production study to identify the relevant prosodic cues of each condition, genitive and dative. The results showed f_0 , pause and duration to be relevant for the indication of a phonological phrase boundary. In a follow-up perception experiment, Gollrad et al. (2010) identified duration to be the most important factor for the disambiguation of syntactic structures in language comprehension.

As the constructions used in Gollrad et al. (2010) are not completely ambiguous and there is also a certain chance of a ‘list intonation’²⁷ with three DPs, the following experiment focusses on fully ambiguous structures containing only two DPs as exemplified in (30). With the dative construction, the phrase boundary would be expected to be placed after the first DP (32a) and after the second DP in the genitive construction (32b).

- (32) a. ... dass der Partner) _{φ} (der Freundin ...
 b. ... dass der Partner der Freundin) _{φ} (...

The aim of the experiment is to identify the different factors (including the less significant ones) contributing to the representation of each case condition. The focus is thus on the factors *given* in a speech signal, and not necessarily the ones that are used by individual listeners to *identify* a specific structure.

3.7.1 Experimental set-up

As briefly stated in Section 3.5, Allbritton et al. (1996) note that subjects will not consistently use prosodic cues to indicate a certain interpretation of syntactically ambiguous sentences, even if the context disambiguates the intended meaning. However, they find that if the speakers were aware of the ambiguity and were asked to pronounce a sentence according to a certain interpretation, the prosodic cues were much more distinct for each condition. In order to control for these findings as well, the experiment involved several types of constructions, which were presented to the participants in the following order:

1. Ambiguous and unambiguous constructions hidden in a larger text, where the ambiguous structures were disambiguated by the context.
2. Unambiguous structures, where the two DPs were masculine (and have distinct determiners as a consequence).
3. Fully ambiguous structures, where the speaker was made aware of the intended meaning.

²⁷List intonation refers to the downstepping intonation pattern used if expressing a list, e.g., as in *I bought an apple, a sausage, an orange and a banana* (see Liberman and Pierrehumbert 1984).

Consequently, full awareness of the ambiguity and a demand for a distinct pronunciation was only given with the sentences of group three. For the experiment, 15 female native speakers of German aged between 20 and 30 were recorded in a sound proof chamber with the data presented on a screen.²⁸ The resulting 432 sentences were hand-annotated syllable-wise using the Praat software (Boersma and Weenink 2013) for the fundamental frequency, the duration of the syllables and for the length of pauses.

3.7.2 Results

Statistical analysis was done with a linear mixed effects regression model (LMER) with subject and item as crossed random factors and the two conditions (genitive and dative) as fixed factors (see also Bögel 2013). The statistical analysis showed the following results:

- A significant drop in fundamental frequency from the alst syllable of the first DP to the determiner of the second DP in the dative condition ($\beta=-9.31$, $SE=2.64$, $t=-3.53$).²⁹
- A significant pause between the first and the second DP in the dative condition: ($\beta=-2.35$, $SE=0.92$, $t=-2.55$).
- The duration of the last syllable of the first DP was significantly longer in the dative condition than in the genitive condition ($\beta=-2.8$, $SE=0.79$, $t=-3.58$).

While these significant patterns are typical indicators of a phonological phrase boundary,³⁰ other prosodic cues are only significant for a smaller subgroup of speakers; i.e., cues are applied individually to a varying extent. For this reason, the statistical analysis was also performed for each individual, because only then can smaller subgroups be identified. Furthermore, it enables the identification of opposing groups of prosodic cues (i.e., where one group of speakers applies a drop in the fundamental frequency, and one group applies a rise at the same position and in the same condition).

Surprisingly, around 33% of the participants did not show any significant results when tested individually for strategies to indicate the dative or the genitive construction (which also shows the importance of a large group of participants). 67% of

²⁸Participants were paid for their efforts.

²⁹ β is the regression coefficient, SE the standard error; the t-value is significant if it is > 2 or < -2 (it replaces the more commonly used p-value within lmer-models).

³⁰There was also a significant larger drop from the second syllable of the second DP to the following verb in the genitive. However, no other indicators for a phonological phrase boundary could be found (no pause, no lengthening of the last syllable) and the data is tricky, as some of the verbs include the unstressed prefix *ge-*, some are particle verbs, where the particle in the first position is stressed as in *zuhören* ‘listen’, and some are verbs without any further prefixes/particles (e.g., *folgen* ‘follow’). As the list of intransitive verbs that allow for an optional dative object is limited, it was not possible to reliably control for this effect as well. This genitive drop thus would have to be confirmed in a separate study and was therefore left out of the current analysis.

the speakers applied linguistic cues to a varying extent; the resulting ‘prototypical constructions’ are shown in the following.

Figure 3.21 represents a ‘prototypical dative’ in Praat, showing four syllables of example (31): *Partner der Freun(din)* ‘partner of/to the friend’. The annotation below gives the reference syllables (non-IPA) and a GToBI annotation, indicating High and Low pitch accents and boundary tones (Grice and Baumann 2002).

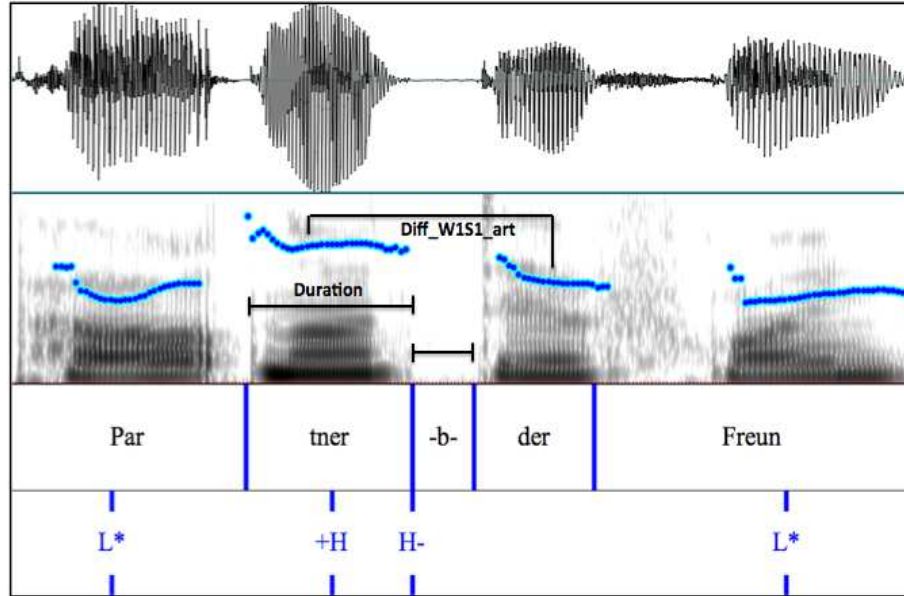


Figure 3.21: A prototypical *dative*.

While the dotted lines in Figure 3.21 represent the fundamental frequency as calculated by Praat, the solid lines have been added by hand and indicate the three most frequently used prosodic cues for the dative construction.³¹

1. **Pause:** 40% of the speakers use a statistically significant pause to indicate a prosodic break between the two DPs.
2. **Duration:** 47% of the speakers lengthen the last syllable of the first DP in a dative construction.
3. **Diff_W1S2_art:** 40% of the speakers have a significant drop in the fundamental frequency from the last syllable of the first DP to the following determiner of the second DP.

While all of these indicators are significant if measured for all participants, they are certainly not true for each individual. Participants used different indicators and

³¹The labels are abbreviations for the specific prosodic cue they indicate. **diff_W1S2_art**, for example, stands for the **d**ifference of the fundamental frequency between the second syllable **S2** of the first noun **W1** and the **a**rticle of the following DP.

combinations thereof to indicate the dative construction. The same is true for the genitive construction given in Figure 3.22.

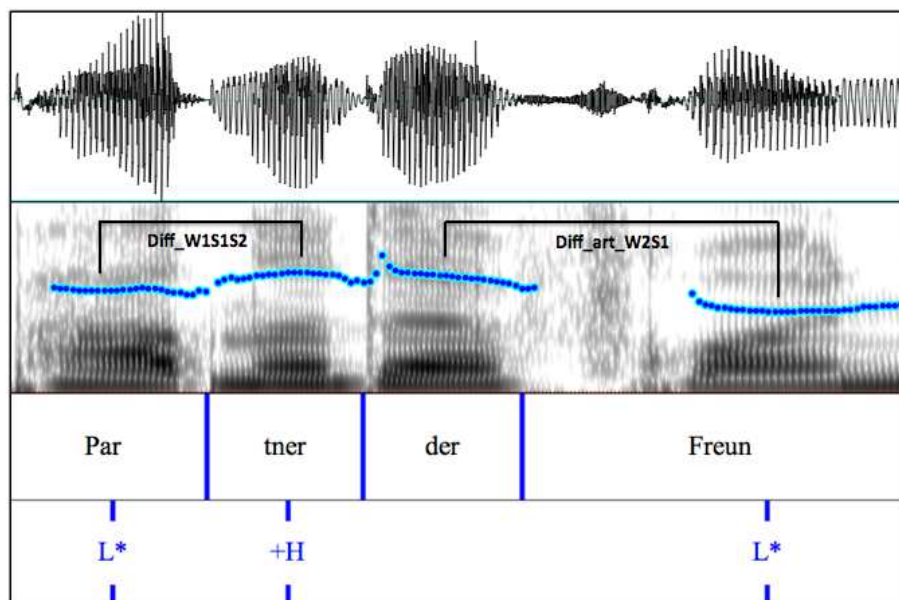


Figure 3.22: A prototypical *genitive*.

First of all, all of the above mentioned (‘dative’) indicators are significant for the genitive as well in that they are *not* present. Indicators visible in the genitive speech signal are

1. **Diff_W1S1S2**: 27% of the participants show a smaller difference in fundamental frequency between the first and the second syllable of the first noun.³²
2. **Diff_art_W2S1**: 20% show a drop from the fundamental frequency value of the determiner *der* to the first syllable of the following noun.

The results that were calculated for each individual showed more prosodic cues, but these failed to satisfy the following constraints which were introduced to ensure a certain significance in the overall strategies on the one hand and to avoid prosodic cues contradicting each other on the other hand. Thus, for a prosodic cue to be considered as a strategic application for the discourse disambiguation, the following constraints had to be met:

1. The phenomenon must be statistically significant for at least 20% of the speakers. As a consequence, prosodic cues that were only applied by, e.g., 10% of the speakers were discarded.

³²It is not clear if the L*+H GToBI annotation is justified here, but it was left in to indicate that there still is a rise in fundamental frequency from the first to the second syllable.

2. Prosodic phenomena that were contradicting each other were excluded as well. That is, for example, if one subgroup of speakers used a High tone on syllable X in the genitive condition and another subgroup used a Low tone on the same syllable X in the genitive condition as well, then this prosodic cue was not considered.

From a machine processing perspective, all of the above described prosodic cues given in a speech signal should be included into the calculation of, e.g., prosodic phrase boundaries, in order to cover as many speakers as possible. Thus, the threshold set by a standard statistical analysis seems to be very high – less frequent cues are excluded from the calculation, even though they might contribute to the recognition of the relevant phrase boundaries. Their lower frequency can be accounted for by a weighted calculation in that more frequent cues are taken to have a strong effect, while less frequent cues have a smaller impact on the overall decision.

Finally, while the statement that speakers do not consistently use one set of prosodic cues (Allbritton et al. 1996) is supported by the findings in this study, there were no significant differences found between the three groups of sentences with growing awareness as discussed above.³³

3.7.3 Discussion and conclusion

The experiment on the dative-genitive alternation introduced above aimed at identifying relevant prosodic cues to disambiguate syntactically ambiguous structures. Following the results given in the previous section, it can be assumed that the prosodic phrasing as it was assumed in (32), repeated below, reflects a specific syntactic interpretation.

- (33) a. **dative:** ... dass der Partner)_φ(der Freundin ...
 b. **genitive** ... dass der Partner der Freundin)_φ(...

The analysis of the data showed that there are several prosodic cues that establish the prosodic phrasing. However, the results have also shown that these cues are not always given; instead, if analysed individually, only about 66% of the speakers show a significant use of prosodic cues that signal a specific syntactic interpretation.

Furthermore, cues should be weighted differently as some are more frequent than others. As a consequence, there is not only a speaker-dependent likelihood for each syntactic construction to be signaled by a phonological phrase boundary, but there is also a likelihood for the prosodic cues that determine such a boundary. While this distinction is not necessarily relevant for theoretical accounts, it can be relevant

³³A discussion of possible reasons, however, goes beyond the scope of this thesis as it does not contribute to the overall aim which was to establish prosodic cues signalling a specific syntactic interpretation.

to know how ‘strong’ a phrase boundary is from a computational perspective as it might be in conflict with calculations at other positions in a sentence.

The following section provides an analysis of the dative-genitive data in LFG, incorporating it into the architecture as it was represented in Figure 3.20.

3.7.4 An analysis of the genitive-dative alternation in LFG

As discussed in Section 3.3.1, from the perspective of comprehension, the speech signal is interpreted linearly syllable by syllable. Each syllable and each pause receives a vector I which encodes the speech signal information, i.e., the different aspects of the speech signal and their respective value at the time the syllable in question is uttered. For example, for the three syllables of the DP ‘der Partner’, the corresponding vectors are given in (34)

$$(34) \quad S: \begin{pmatrix} \text{segments} \\ f_0 \\ \text{duration} \\ \dots \end{pmatrix} (n) \quad \text{would yield} \quad \begin{pmatrix} [de^v] \\ 192 \text{ Hz} \\ 0.15 \text{ s} \\ \dots \end{pmatrix} (1) \quad \begin{pmatrix} [pa^v] \\ 181 \text{ Hz} \\ 0.25 \text{ s} \\ \dots \end{pmatrix} (2) \quad \begin{pmatrix} [tn\emptyset] \\ 269 \text{ Hz} \\ 0.17 \text{ s} \\ \dots \end{pmatrix} (3)$$

These vectors are merged into the p-diagram (Figure 3.23), enabling a fine-grained representation of the original speech signal. Once encoded in the p-diagram, the information can be easily extracted; for example, the function $(S_2 \text{ DURATION}) = 0.25 \text{ s}$ refers to the second vector’s value for the attribute DURATION which equals 0.25 seconds.

↑	...	↑	↑
PHRASING	-	-	∅	-	-	-	
SEMIT_DIFF	...	-1	6.8	-4.3	-1.9	2.6	
GTOBI	-	L*	+H H-	-	L*	+H	
BREAK_IND	-	-	3	-	-	-	
DURATION	0.15	0.25	0.25	0.13	0.31	0.19	
FUND._FREQ.	192	181	269	209	188	218	
SEGMENTS	[de ^v]	[pa ^v]	[tn∅]	[de ^v]	[fʁʊçən]	[dɪn]	
VECTORINDEX	S₁	S₂	S₃	S₄	S₅	S₆	

Figure 3.23: P-diagram for the speech signal given in Figure 3.21.

In addition to the speech signal information, Figure 3.23 also includes an interpretation section which, on the basis of the raw data in the speech signal, interprets

the information in a more abstract way. Part of this interpretation is also the calculation of a phonological phrase boundary for which several factors can be taken into account: a) The strong rise in pitch to the otherwise unstressed second syllable of ‘partner’ and the following drop (SEMIT_DIFF) indicate a high boundary tone (GTOBI), b) the pause after ‘partner’ (BREAK_IND), and c) the relatively long duration of the unstressed second syllable of ‘partner’ (DURATION). The resulting phonological phrase boundary is one relevant factor at the prosody–syntax interface, as it matches and thus signals a specific syntactic phrasing during the *transfer of structure*.

The other factor to the interface communication is the multidimensional lexicon, which allows for an alignment of speech signal segments with their corresponding syntactic form during the *transfer of vocabulary*.

concept	s-form	p-form
PARTNER	S-FORM <Partner>	P-FORM [ˈpa ^h .tne]
	N (↑ PRED) = ‘Partner’	SEGMENTS /p a ^h t n e/
	(↑ NUM) = sg	METRICAL FRAME (‘σσ) _ω
	...	
ARTICLE	S-FORM <der>	P-FORM [de ^v]
	D (↑ PRED) = ‘der’	SEGMENTS /d e ^v /
	(↑ NUM) = sg	METRICAL FRAME σ
	(↑ GEND) = fem	
	(↑ CASE) = {gen dat}	
...		

Table 3.5: Lexical entries for *der* and *Partner*.

The previous discussion on prosodic cues and their influence on syntactic phrasing (Section 3.6 and 3.7.2) came to the conclusion that hard constraints³⁴ need to be replaced by softer constraints to allow for more flexibility. One possibility to express probabilities in c-structure rule annotations is discussed in the following section.

3.7.5 Disambiguating c-structure via ranked constraints

As has been shown in Section 3.7.2, the prosodic indications are numerous, but ranked by frequency in that some of them are applied by almost 50% of the speakers, while others are only used by 20% of the speakers. As a consequence, prosodic cues cannot be taken to be ‘hard constraints’ which either allows or prohibits a certain analysis.

For these specific cases, the grammar writing platform XLE (Crouch et al. 2015) allows for OT-like constraints in the c-structure rule annotation as originally proposed by Frank et al. (1998) and extended and modified in the current XLE docu-

³⁴E.g., as they were used in the case of the c-structure rule annotation for coordination in (27).

mentation (Crouch et al. 2015). ‘OT-like’ in this context means that the notion of constraints is not understood as in general Optimality Theory (Prince and Smolensky 2004) or in OT-LFG (Bresnan 2000) (see also Chapter 4, Section 4.3.1). The underlying assumption is not an infinite or a slightly restricted set of candidates that are ranked according to the OT-like annotations. Instead, these annotations are mostly added to fully-constrained grammars that nevertheless allow for more than one analysis of an input candidate.

The OT-like marks given in a c-structure annotation thus rank the different syntactically correct analyses of one common input candidate: A preference/dispreference mark can be added to a specific annotation based on further information, which can come from other modules, but might also be information known to the researcher to exist outside of the scope of grammar, e.g., the frequency of a specific analysis. Also, OT only allows for negative constraints ruling out impossible/unlikely candidates, but XLE also enables the user to mark a certain condition as being preferred via so-called preference marks. Dispreference marks coincide with the original idea of OT and are generally used on rare, yet grammatical constructions. Preference marks on the other hand are unique to the XLE implementation and are applied if one reading is preferred.³⁵

This system enables the user to soften the standard ‘hard’ constraints provided by XLE and allows for the implementation of phenomena, whose analysis cannot be easily divided into ‘good’ and ‘bad’ as it is the case with the disambiguation of the genitive-dative alternation by means of prosodic cues. That is, the constraint that a syntactic dative construction only applies if there is a phonological phrase boundary after the first DP cannot be analysed as a hard constraint, but must be implemented as a soft constraint via OT-like constraints.

As discussed above in Section 3.7.3, the prosodic cues which lead to the assumption that a phonological phrase boundary is present are also ranked among themselves according to frequency, indicating a strong or weak phonological phrase boundary. This results in a twofold choice: Either the singular prosodic cues are weighted within a c-structure annotation, thus indicating also the strength of that boundary, or the phrase boundary is calculated solely p-structure internally and its ‘strength’ is taken to be irrelevant to syntax; that is, only the information that a phrase boundary is present is available at the interface.

The former approach was pursued in Bögel (2013) which led to c-structure annotation constraints similar to the following:

- ‘If semitone difference between first and second DP greater or equal to 2, then select dative (SEMIT_DIFF \geq 2) with a probability of 40%’.

³⁵Note that both methods have problematic issues in that they can lead to preference/dispreference analysis between completely unrelated constructions (for more on this topic see Frank et al. 1998). Thus, an implementation with OT marks should always carefully consider other possible, but unintended, interactions.

- ‘If pause after first DP equals 3 ($\text{BREAK_IND} = 3$), then select dative with a probability of 40%’.
- ‘If semitone difference between the first and second syllable of the first DP is smaller than 2 ($\text{SEMIT_DIFF} < 2$), select the genitive with a probability of 20%’.

The above constructions provide an internal ranking determined by the probabilities with which each prosodic cue appears. However, this approach has one important drawback: It is not modular! Instead c-structure refers to attributes that are clearly part of p-structure/the speech signal. The sole reference to phonological phrase boundaries, on the other hand, is a reference to the only material available at the interface between syntactic and prosodic structure as it was also discussed in Chapter 2. Prosodic constituency is seen as an abstract representation determined by p-structure internal prosodic cues, which signals the prosodic phrasing structure at the prosody-syntax interface. As a consequence, the c-structure annotation below will only relate to prosodic constituency.³⁶

As was established in Section 3.5, the influence of prosody on syntax is limited: If in direct competition, the syntactic phrasing will always be preferred over the prosodic phrasing. Prosodic phrasing thus supports syntactic phrasing, but it only determines syntactic phrasing in cases of complete ambiguities. For this reason, prosodic phrasing does not ‘project’ to syntactic phrasing in the approach presented in this thesis. Instead, syntactic phrasing ‘checks’ for a specific prosodic phrasing via the composition of the inverse relation of π with ρ : $\rho(\pi^{-1}(f))$, abbreviated as $\natural(f)$. This relation allows c-structure annotation rules to refer to information, which is relevant for a specific c-structure analysis, but which is encoded in p-structure.

In (35), an (XLE-)implementation of the phrase structure rule annotation referring to the dative construction is shown.

$$(35) \text{ Dative: } VP \rightarrow \begin{array}{c} \text{DP} \qquad \qquad \text{DP} \qquad \text{V} \\ \{ (\natural(T(*))) S_{max} \text{ PHRASING} \} =_c \varphi \\ \text{PHPBREAK} \in \mathfrak{o}^* \\ |(\natural(T(*))) S_{max} \text{ PHRASING} \} \neq \varphi \} \end{array}$$

The $\natural(T(*))$ in the first conjunct of (35) refers to the (set of) terminal nodes connected with the DP and the projection between c-structure and p-structure as described in Section 3.5. $(S_{max} \text{ PHRASING}) =_c \varphi$ is a reference to the last vector of the set of DP terminal nodes (i.e., the one with the maximum value, which is the last syllable of the first DP). The dimension PHRASING in that vector must yield the value φ , indicating a phonological phrase boundary. If this is the case, then the

³⁶Note that this restriction to prosodic constituency might not be true for the interfaces of other modules with p-structure. For example, it can be assumed that information structure has an interest in pitch, i.e., in high and low tones.

construction receives a preference mark PHPBREAK, indicated by $\text{PHPBREAK} \in \text{o}^*$, which translates as ‘the mark PHPBREAK is element of the o(ptimality)-structure’. If PHRASING does not yield φ , the c-structure rule is still valid, as the second conjunct in (35) allows for this to be the case ($\neq \varphi$). However, if there is another analysis in competition and this analysis carries a preference mark, then the default analysis is overruled.³⁷

The OT-like marker PHPBREAK is evaluated according to a preestablished ranking as given in (36), which shows the optimality order via preference marks, indicated by +.³⁸

(36) OPTIMALITYORDER +PHPBREAK

The application of OT-like soft constraints to the c-structure annotation rules allows for some flexibility when dealing with concrete speech signal data, as it only indicates a preference when a specific information is present, but does not automatically prohibit the rule from firing, if the relevant information is *not* present. OT-like constraints thus enable the implementation of performance factors, for example the frequency of a specific prosodic cue used by a group of speakers. In that sense, OT-like constraints are the pivot between real-case performance results and the rules and constraints of the ‘core’ grammar.

In cases where the information necessary for the disambiguation of syntactically ambiguous structures cannot be discerned from the data (i.e., from p-structure), the c-structure annotation given above still allows for the unmarked case to be parsed. However, if preference marks are given with a *competing* analysis, the unmarked analysis is discarded. If, on the other hand, no preference marks apply with both competing analyses, the unmarked analysis is equal to other unmarked analyses of the ambiguous input, which is justified given the non-existent preference for one specific analysis in the speech signal.

3.8 Summary and conclusion

The aim of this chapter was two-fold: First, a new way of representing prosodic information was introduced: the p-diagram. The p-diagram allows for all relevant

³⁷For a more elaborate interaction between preference marks from a machine processing perspective, see Frank et al. (1998) in general, and for this specific problem Bögel (2013), which is, however, not modular.

³⁸The representation is as it would appear in the XLE grammar configuration section. In the case presented here, it does not really matter if preference or dispreference marks are used. Preference marks, which are not in line with classical OT, were chosen, because they save an addition of the prosodically unmarked case (i.e., an expression which has neither distinct prosodic dative nor genitive cues) to the optimality order as the most dispreferred case. This would be necessary because otherwise the unmarked case would always be the preferred case. Here, the unmarked case would only be the preferred analysis, if none of the constructions with preference marks apply. An extra marking of the unmarked case is thus unnecessary.

elements of the speech signal to be encoded in a compact and easily accessible representation. Its syllable-based foundation enables a representation independently of the (rather instable) formation of prosodic constituency. Furthermore, the representation as such is theory-independent, as it is left to the researcher how the data is interpreted at the level of interpretation.

Second, this chapter established the relation between c- and p-structure. This correspondence is assumed to apply on two levels: a) the *transfer of vocabulary*, which aligns the segments given in a speech signal with the phonological representations stored in the multi-dimensional lexicon. The multi-dimensional lexicon then associates the phonological forms extracted from the speech signal with the related syntactic forms which are the basic foundation of syntactic analysis. b) the *transfer of structure* allows for a direct and vocabulary-independent correspondence between c- and p-structure which communicates information on prosodic constituency, i.e., on structure, present in a given speech signal.

This twofold correspondence, structural and lexical, was exemplified by the analysis of two concrete language phenomena: English coordination grouping as it was reported by Wagner (2005) and an experiment on German dative/genitive alternation, where the results also showed that the influence of prosodic cues on syntactic phrasing is limited, in that only about two thirds of the speakers used discernable prosodic cues to communicate the intended interpretation.

This chapter was written from the perspective of comprehension (from form to meaning): Concrete prosodic cues that indicate the intended interpretation of fully ambiguous structures were determined and analysed and a first version of the interface was established on the basis of the findings. The following chapters are written with a focus on the production perspective and will establish a more fine-grained view on the prosody–syntax interface and the related components. While Chapters 5 and 6 focus on the transfer of structure and on resulting conclusions about the interface architecture, the following Chapter 4 discusses the transfer of vocabulary in more detail and introduces the p–structure internal processes of postlexical phonology.

Chapter 4

Postlexical phonology and the syntax-prosody interface

4.1 Introduction

The previous chapters introduced the multidimensional lexicon and the p-diagram and formalized the prosody-syntax interface from the perspective of syntactic ambiguities. The focus of the discussion was on the comprehensive perspective on grammar, i.e., how the processing of a sound can influence a syntactic decision (and consequently the meaning of a string). A concrete example of how a given speech signal can be encoded in p-structure and is made available at the interface to assist in the disambiguation of syntactic ambiguities was given in Chapter 3, Section 3.7.4.

The following chapters are concerned with the ‘opposite’ direction: How meaning can be transferred to sound, whereby the discussion is (naturally) restricted to the transfer of information from the lexicon and c-structure to p-structure, i.e., the syntax→prosody interface. As briefly introduced in the previous chapter, this transfer is assumed to be twofold: On the one hand, *vocabulary* needs to be transferred, i.e., the s-forms of the terminal nodes of the c-structure tree have to be ‘translated’ into the corresponding p-forms via the multidimensional lexicon. The lexical phonological information encoded along with each p-form thereby becomes available to further postlexical phonological processes. On the other hand, *structure* has to be transferred as well, creating basic prosodic units from syntactic (and lexical) structures which are then adjusted according to a set of postlexical principles stipulating domains within which a specific postlexical phonological rule can apply. Thus, in this thesis, *structure* refers to syntactic and prosodic phrasing above the word level, while *vocabulary* refers to the information that is stored in the lexicon along with each phonological form.

This chapter will provide a more detailed introduction into the *transfer of vocabulary* in the case of the first person singular nominative (1SgNom) pronoun in Swabian which distinguishes between a weak (clitic) and a full form in medial clause position.

Depending on the information structure status, the respective lexical entry is realised and the corresponding phonological information is transferred to p-structure.

In addition to the transfer of (lexical) vocabulary, this chapter also focusses on the application of postlexical phonological processes and shows how the raw (and isomorph) data transfer from lexicon to p-structure is converted step by step into the final output of p-structure.

The present chapter is structured as follows: First, a short introduction into two theoretical concepts relevant for the following two chapters is given, namely clitics and postlexical phonology. The introduction will also motivate the underlying assumptions that determine the architectural decisions made in this chapter. Section 4.2 introduces data on the different realizations of the Swabian 1SgNom pronoun and shows where each of them originates, thus distinguishing between different modules relevant for the generation of each variant. Following this is a discussion on the postlexical phonological rules involved in the generation of possible surface forms for a given Swabian clitic cluster (Section 4.2.3). Section 4.2.4 then describes a concrete analysis of the theoretical findings within LFG. The chapter concludes with an introduction to *Optimality Theory* in general and in particular within LFG and discusses why OT is ‘unoptimal’ from the perspective of this thesis.

While the focus of this chapter is on the transfer of information from the lexicon to p-structure (*transfer of vocabulary*), the following Chapters 5 and 6 will extend the discussion (and the interface processes) with further data and show how *structure* above the level of the prosodic word is transferred from syntax to p-structure.

4.1.1 On clitics

The following chapters analyse a number of clitic phenomena, namely Swabian 1Sg-Nom pronoun clitics and Degema and Pashto endoclitics (Chapter 5 and 6). It is thus necessary to quickly introduce the assumed theory of cliticization, i.e.: how is a clitic defined in this thesis?

The attempt at defining clitics in a typological manner and to find common criteria to distinguish clitics, affixes, and words continues to be notoriously difficult. In a first analysis, Zwicky (1977) distinguishes between three types of clitics: simple clitics, special clitics and bound words. Special clitics are characterized as unaccented bound forms that have corresponding full/free forms and are subject to ‘special’ syntactic rules. Simple clitics on the other hand are phonologically reduced variants of free forms that display the same syntactic distribution as their free counterparts. Bound words, as a third category, are unaccented and do not have a corresponding free form. Thus, in this first proposal, the following two features are important for the distinction between various types of clitics: The respective syntactic distribution and the relationship between the clitic and its free counterpart (if present). There is no distinction concerning stress: All clitic types are considered to be unaccented and prosodically deficient.

While the separation of simple and special clitics was retained in the subsequent

literature and the terms are still frequently used, the notion of the bound word was abolished (Anderson 2005, 12, among others).¹ In his book on clitics, Anderson (2005) notes that the existence of a free/full variant is not an important criterion for the classification of clitics. He furthermore states that the phonological derivation between a clitic and its full form is not always given (a view supported by Spencer and Luís (2012)); thus, Anderson does not take the relationship of a clitic to its corresponding full form (if present) into account when determining the different types of clitics. Instead he proposes three types of clitics on the basis of two characteristics: 1. clitics which are prosodically deficient, 2. clitics which act according to idiosyncratic syntactic rules (in comparison to other elements in the language)² and 3. clitics which display both characteristics. Clitics of type 2 and 3 are usually treated under the term *special clitics*, while clitics of type 1 describe the former *simple clitics*.

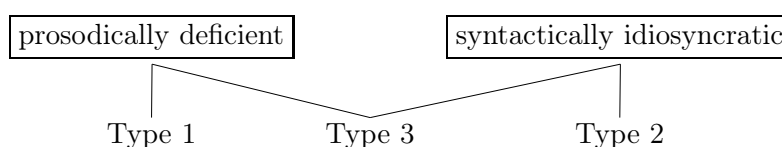


Figure 4.1: The three-way distinction of clitics proposed in Anderson (2005).

This thesis follows the distinction of clitic types proposed in Anderson (2005) and supports his assumption that clitics are ordinary lexical items that form independent terminal nodes in the syntactic tree, but are (in most cases) prosodically deficient and thus have to be phonologically attached to their respective host during the postlexical determination.³

Another important aspect of the theory of clitics is the distinction between clitic and affix which has been the topic of uncounted literature since the first (and up to now very influential) set of criteria was proposed by Zwicky and Pullum (1983). In the following, these criteria are listed and discussed.

CRITERION A: Clitics can exhibit a low degree of selection with respect to their hosts, while affixes exhibit a high degree of selection with respect to their stems.

¹Note also that the line between ‘simple’ and ‘special’ clitics concerning their syntactic distribution is not always easy to draw, as Spencer and Luís (2012, 96ff.) demonstrate with a number of examples from different languages: While a clitic might mostly stay conform to the syntactic distribution of its full form, it might vary in unexpected ways. However, as the distinction of various types of clitics is not in the focus of the thesis, this topic will be left for further research.

²Note that this group is in principle able to bear stress, thus abolishing the understanding of clitics as principally unstressed elements.

³While the syntactic distribution is straightforward for the notion of simple clitics, special clitics (especially endoclititics and second position clitics) are problematic in this context. However, these issues can be resolved as well. As the discussion would lead too far at this point, an introduction to these problematic cases (and their solution) is postponed to Chapter 5 and 6, respectively.

Affixes usually are restricted to one particular word class or even a subgroup of this category. Clitics on the other hand seemingly do not necessarily depend on a specific word class to attach to. However, this decision is also determined by the position of the clitic: If the clitic has a fixed position within a category-internal construction, as it is for example the case with Urdu *Ezafe* (cf. Bögel and Butt 2013), then this criterion might not apply. It is thus important to understand that not all of these criteria need to apply, but that they are indicators which may or may not be true for a specific clitic.

CRITERION B: Arbitrary gaps in the set of combinations are more characteristic of affixed words than of clitic groups.

Criterion B essentially states the fact that there are often gaps in an affixal paradigm, but that gaps are not common with clitics – an observation which is in line with the assumption that clitics are individual syntactic items and subject to postlexical phonological rules which are considered to be ‘exceptionless’, see 4.1.2.

CRITERION C: Morphophonological idiosyncrasies are more characteristic of affixed words than of clitic groups.

CRITERION D: Semantic idiosyncrasies are more characteristic of affixed words than of clitic groups

Criteria C and D express similar concepts: A clitic is not supposed to show idiosyncrasies triggered by host-internal properties, neither in its structure (phonology/morphology) nor in its meaning.

CRITERION E: Syntactic rules can affect affixed words, but cannot affect clitic groups.

This criterion states that clitics and their hosts are not treated as a syntactic unit in contrast to affixed words, e.g., a clitic might take scope over a coordination, but an affix has to be attached to each conjunct.⁴

CRITERION F: Clitics can attach to material already containing clitics but affixes cannot.

This refers again to the distinction between lexical and postlexical processes. Affixes are attached lexically; thus, once clitics are attached postlexically, affixation is completed.

As has been already indicated in the previous discussion, the criteria to distinguish clitics and affixes are not clear-cut and do not necessarily have to apply as a whole. However, it is assumed in this thesis that at least a subset of these criteria should apply. The following chapters will introduce a wide variety of clitic types:

⁴But see, e.g., Kabak (2007) for a concept on ‘suspended’ affixation scoping over coordinated structures.

simple clitics (Swabian) which have a corresponding full form whose syntactic distribution they share, special (endo)clitics (Degema) whose realization is determined by phrase boundaries and which do not have a full form counterpart, and a subset of the special clitics, the second position (endo)clitics found in Pashto. However, it needs to be stressed that this thesis does not aim to establish a specific theory of clitics — clitics just ‘happen to be’ interesting cases positioned between the modules and, as a consequence, they shed light on the respective processes at the interfaces and beyond. They are thus a means to an (architectural) end.

4.1.2 Postlexical phonological rules

As the name already suggests, postlexical phonological rules are those rules which apply *after* the lexicon in contrast to the lexical phonological rules which are assumed to interact with morphology in the theory of lexical phonology (Mohanan 1982, Kiparsky 1982b). This distinction is closely related to the *Strong Lexicalist Hypothesis* (Halle 1973) which assumes that all word formation, derivational and inflectional, is part of the lexicon as opposed to the larger constructions above the word level which are formed by syntactic rules.⁵ The idea that each of these components, lexical and postlexical, has a separate set of phonological rules is first most clearly expressed in Rubach (1981), Mohanan (1982) and Kiparsky (1982b) where the authors show that some phonological rules apply below the word level and some above the word level.

This chapter focusses on postlexical phonological rules which are assumed to apply *after* the lexical item has been fully formed in the lexicon and after the resulting words have formed a string. At this point, evidence on internal morphological word structure is no longer accessible to the postlexical phonological rules, which is also in line with the *Opacity Principle* (Mohanan 1982, p.23). According to the *Opacity Principle*, the internal morphological structure of one level is invisible to the processes at another; thus, the internal structure of a word in a given string is no longer accessible to postlexical (syntactic OR phonological) rules. Therefore, postlexical phonological rules are triggered solely by the phonological structures as they are present in the string⁶ and can apply across word boundaries as well as word internally; they might even introduce new material which was not part of the underlying phonological inventory (see, e.g., Lahiri (2000, 171) for an example in Dutch). As a consequence, postlexical phonological rules are exceptionless: If a

⁵This is also one of the fundamental assumptions of Lexical Functional Grammar expressed in the *Principle of Lexical Integrity*, see Asudeh et al. (2013) for a recent discussion and Chapter 2, Section 2.3.2 for a brief introduction to the concept with respect to LFG.

⁶Note, however, that this is most probably also true for lexical phonological rules which are only superficially triggered by a morpheme added to the material at a certain level. In reality, however, they are triggered by, e.g., changed metrical structures, which is conform with the modular assumption where phonology will only react to phonological triggers, and morphology to morphological triggers, see Lahiri (2000, 168) for further discussion.

specific phonological trigger is given in a language, then the corresponding rule will always fire.

The scope of most postlexical phonological rules is limited by prosodic domains, as many postlexical processes are bound to a specific level of the prosodic hierarchy. Consider, for example, the phenomenon *Tuscan gorgio* found in the Tuscan variety of Italian, which requires the voiceless stops /p, t, k/ to be pronounced as the corresponding fricatives ([ɸ, θ, h]) between two vowels: /la kasa/ is thus realized as [laha:sa]. However, while this process applies within the prosodic domain of the intonational phrase, it cannot cross the boundary between two subsequent intonational phrases and is thus an example for a sandhi rule restricted to this domain.⁷

The interaction between specific sandhi phenomena and prosodic domains requires some pre-planning and supports the assumption that postlexical rules are non-cyclic in nature. That is, they do not reapply each time after a word has been added to the linear string or, alternatively, from the innermost syntactic constituent to the outside, but they apply to the string as a whole (or possibly in pre-planned larger prosodic domains) and are thus not interspersed with syntactic processes (Kiparsky 1982a).

The resulting architecture proposed by, e.g., Kiparsky (1982b, 4) assumes that the lexically fully (phonologically and morphologically) formed words are inserted into the syntactic structure provided by the syntactic component and that the postlexical phonological rules operate on the resulting syntactic string.

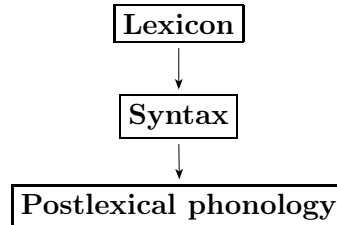


Figure 4.2: Architecture generally assumed in (post-)lexical phonology.

While this representation fits in well with an architecture of language in the tradition of the ‘T-model’ (Chomsky 1981, see also Chapter 2, Section 2.4.2), it is not clear what exactly is inserted into the syntactic structure: Is it the complete lexical item, with all of its syntactic, semantic and phonological aspects? How is the phonological information encoded so it is available to the postlexical phonological rules, e.g., how are stress patterns and prosodic status transferred? And, specifically in transformational grammars, how are abstract terminal nodes encoding past tense or inflection taken into consideration with respect to the resulting phonological effects?

As stated in the previous chapter, this dissertation proposes a slightly different model. It assumes the multidimensional lexicon as it has been introduced in Chapter 3. Following Levelt et al. (1999), the multidimensional lexicon distinguishes

⁷See Nespor and Vogel (1986, 205ff) for further details on the phenomenon.

between a syntactic form, a phonological form and a concept in the lexical representation of each lexical item, thus also preserving the modularity between the different components in the lexicon. Under this approach, the lexicon can be viewed as a reference book between the syntactic (c-structure) and the postlexical phonological (p-structure) component, as shown in Figure 4.3.

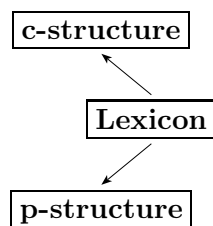


Figure 4.3: The position of the lexicon as proposed in this thesis.

This representation is more in line with the proposals that assume a parallel but modular architecture between the components (e.g., Jackendoff (2002), this thesis), where each module has its own set of principles and constraints and is built in parallel with reference to the other components by means of, e.g., the multidimensional lexical entries. Postlexical phonology is part of the p-structure component and is applied to information transferred from the lexicon (and from syntax, Chapter 5 and 6) thus generating the ‘surface’ output of the p-structure component.

In the following section, the transfer of lexical information to p-structure and the subsequent application of postlexical phonological rules is introduced via the Swabian first person singular nominative pronoun alternation.

4.2 Case study: Swabian pronoun alternation

Swabian, a southern German dialect spoken by approximately 800,000 speakers, shows an allophonic variation for the first person singular nominative (1SgNom) pronoun in medial clause position.⁸ In the following example ((37)), the personal pronoun is realized as an unstressed [ə] in the matrix clause and in its fully stressed form [i:] in the subordinate clause.

- (37) jɛtst kəx=ə ɛbəs vo: blo:s ?i: kɛn
Now cook.1SG.PRS=1SG.NOM something of.which just 1SG.NOM know.1SG.PRS
Now I will cook something of which just I know.’

While the matrix pronoun is not in the focus,⁹ the pronoun of the subordinate clause is in the focus of the clause, also indicated by the use of the intensifying

⁸Where medial clause position refers to the position after the finite verb (+particle) or the complementizer. Clause-initial and clause-final pronouns are left out to keep the data as compact as possible.

⁹Focus in general expresses the ‘most important part’ or what is ‘new’ in the utterance – however, Krifka (2008) notes that these notions are sometimes difficult to apply and that they should rather

focus-sensitive particle [blo:s] (‘only’); i.e., the allomorphic variant at the medial clause position is determined by the information structure status of the pronoun.

This difference is also expressed in terms of prosodic status: As will be explained in more detail in Section 4.2.2.2, the two variants behave differently with respect to the preceding element. While the host preceding the unstressed (clitic) pronoun resyllabifies to the following [ə] indicating a shared prosodic word domain, the focussed pronoun [i:] receives independent prosodic word status, indicated by a non-present resyllabification of the previous word’s final consonant to the onset of the pronoun and a glottal stop [ʔi:] before the stressed vowel in context.

One option of analysing the distribution of the clitic vs. the full form is the application of postlexical phonological rules which would derive the cliticized version of the pronoun from the corresponding lexical full form. However, in the above alternation, there are no specific *phonological* environments explaining the alternation between [i:] and [ə] (see also Young-Scholten 1993, 47ff), but the distribution is determined by conditions given in information structure (focus). As a consequence, both phonetic variants of the pronoun will have to be listed in the lexicon.¹⁰

However, there is a third alternation to the 1SgNom pronoun which is determined by the postlexical environment of the clitic and consequently does not constitute a separate lexical p-form, but is derived from the clitic pronoun via the application of postlexical phonological rules: The subject drop.

4.2.1 A special case: 1SgNom pronoun subject-drop

Standard German is not a pro-drop language. However, some of the German dialects allow for some version of pro-drop, among them Swabian. While the second person subject drop is applied very frequently and is mainly determined by the pronoun’s syntactic position (cf. Haag-Merz 1996), the first person optional subject drop is determined by postlexical phonological conditions and is used rather seldomly.¹¹

- (38) Gɛftən hən(=ə)=də b̥aɪ dɐ bɔft gɛvə
 Yesterday have.1SG.PRS(=1SG.NOM)=2SG.ACC at the post.office see.PRF
 ‘Yesterday, I saw you at the post office.’

be seen as statistical correlations, but not as definitional features. Instead Krifka comprises a wide variety of focus types under the definition of focus as ‘the presence of alternatives that are relevant for the interpretation of linguistic expressions’ (Krifka 2008, 247), i.e., the possibility of alternatives indicates covert questions suggested by the context about the focussed element(s) (e.g., ‘Who knows the recipe?’ – ‘Only I do’). I follow this broad definition of focus for the Swabian data presented here and leave the more fine-grained focus definition of the distinction between the Swabian pronoun types to further (information-structural) research.

¹⁰For a discussion on the lexical status of clitics, see Anderson (2005, 27).

¹¹Ca. 1.3% of all first person subject pronouns, see Bohnacker (2013). This is also owed to the very specific environment needed for the first person subject to be dropped, which is described in the following section.

In her book on Swabian pronouns, Haag-Merz (1996) defines concrete conditions for the subject-drop in the first person. In order to narrow down valid phonological processes concerning the Swabian 1SgNom pronoun, these assumptions are compared to my own dialect version¹² and to the findings in Bohnacker (2013) in the following.

According to Haag-Merz (1996), the subject-drop can only occur, if the corresponding form would be the unstressed pronoun (the clitic [ə]). This constraint can be confirmed in my own dialect version.¹³

Another condition raised by Haag-Merz is that the dropped subject pronoun must be part of a clitic cluster. This rule finds only one exception in sentences, where the speaker inserts [gləʊb] ('I think') in medial clause position.

- (39) de:ɐ hat gləʊb ø gvõnə
 3SG.M.NOM have.3SG.PRS think.1SG.PRS (1SG.NOM) win.PRF
 '(I) think, he won.'

In this position, [gləʊb] is used as 'a discourse-marker, a parenthetical *I think* comment' which is 'near-formulaic' (Bohnacker 2013, 277). As this combination of [gləʊb] and the optional subject-drop has an exceptional (grammaticalised) status in the discourse and can be easily covered by an individual constraint, the clitic cluster as the standard environment for the subject drop is still regarded as a valid constraint.

Following Nübling (1992), Haag-Merz (1996) claims that the ultimate syllable of the host has to be stressed. The underlying assumption is that the following subject clitic is then further destressed and can be deleted. Bohnacker (2013), on the other hand, shows an example, where this constraint does not apply.

- (40) ... , ε'r ənɐ(=ə)=mə drɐ, ...
 remember.1SG.PRS(=1SG.NOM)=1SG.ACC at.it
 '(I) remember it/(remind me of it).'
- (modified from Bohnacker 2013, 275, ex.29)

In (40), lexical stress is on the second syllable, followed by two unstressed syllables before the first pronoun. Note that in case of a subject drop the example in (40) is irritating because it could also mean 'remind me (of it)', i.e., the imperative form. This ambiguity, caused by the subject drop, is probably the reason that such constructions with finite verbs before the clitic cluster are usually avoided – rather, the spoken dialect prefers the use of perfective constructions, where the finite auxiliary is always monosyllabic. As a consequence, the last syllable is almost always stressed before the subject clitic.

¹²The Swabian dialect version spoken by myself is from the region around the city of Göppingen, but also influenced by the Swabian spoken in Stuttgart, as both parents grew up there.

¹³It is, however, hard to judge the findings with respect to the unstressed pronoun in Bohnacker (2013), as she does not distinguish between [i:] and [ə], but uses the form [i:] in all of her examples.

- (44) ? Gɛftən hən=∅=m k^hɔlfə
 Yesterday have.1SG.PRS(=1SG.NOM)=3SG.M.DAT help.PRF
 ‘Yesterday, I helped him.’

In this case, the situation is slightly more complex: The subject clitic can (and not ‘must’) be deleted, if the syllabified consonant is retained ([hən(=ə)=m]). Another possibility is the consonant’s loss of syllabic status: In this case, the consonant forms part of the subject clitic’s coda ([hən=ə=m]). As a consequence, the subject clitic has to be retained in order to avoid a stray consonant [m] (i.e., an isolated consonant not grouped into a syllable). It can thus be concluded that the subject clitic does not have to be dropped obligatorily, but that there are several options to this construction which are restricted by a constraint to preserve a valid syllable structure. The following constructions support this restriction.

Up to now, there are two confirmed conditions for an optional subject drop: First, the subject pronoun needs to be unstressed, and second, it needs to be part of a clitic cluster. Example ((45)) shows a phonological environment in which the above conditions are met, but where the subject clitic must not be dropped.

- (45) *Gɛftən hən=∅=s ufɣmaxt
 Yesterday have.1SG.PRS(=1SG.NOM)=3SG.N.DAT open.PRF
 ‘Yesterday, (I) opened it.’

However, compare (45) with (46), where the object clitic [s] is followed by a further clitic, the syllabified [ŋ] (a particle, *denn* in Standard German), and (47), where the subject is in the clause-initial position and the object clitic [s] is not contained in a clitic cluster.

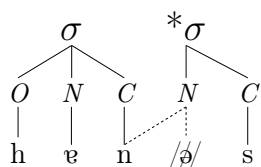
- (46) hap=∅=s=ŋ nō drɪn
 have.1SG.PRS(=1SG.NOM)=3SG.N.ACC=then still in.it
 (The speaker is looking for something) ‘Do I still have it in here then?’
 (Bohnacker 2013, 275,ex.30)

- (47) i: hən=s ɣsɛɐ
 1SG.NOM have.1SG.PRS=3SG.N.ACC see.PRF
 ‘I saw it.’

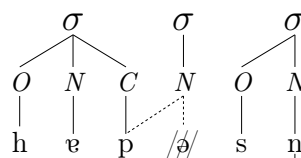
The question here is why (45) is ungrammatical in comparison to (46) and (47), especially as (45) and (47) result in a common surface form ([həns]). Seen from a *phonology-only* perspective, the reason for this is a constraint that requires a valid syllable structure. This is demonstrated in Figure 4.4, where it is assumed

- that the subject drop occurs after syllabification,
- that each syllable needs to consist of at least a nucleus,

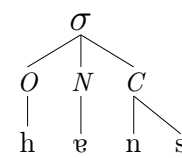
- that syllabified consonants (marked χ) inhabit a nucleus, and
- that each segment has to be grouped with a syllable.¹⁴



Example (45)



Example (46)



Example (47)

Figure 4.4: Syllabification, subject drop, and invalid syllable structures.

In (45), during the syllabification process (and before the subject drop), [s] forms the coda of the second syllable, with the subject clitic [ə] as its nucleus. However, if the subject clitic is dropped, the [s] is left astray, violating the notion that all segments must be part of valid syllables.¹⁵ In example (46), on the contrary, [s] is grouped together with the following syllabified η during the syllabification process and inhabits the onset position obeying the principle of onset maximisation. In the case of a subject drop, no consonant is left astray and the valid syllable structure constraint is not violated. In (47), finally, [s] is phrased into one syllable with the preceding auxiliary [hən]. While the surface form may look similar to the one of (45), a look at the different syllable structures makes it obvious, why (45) fails, while (47) is valid.

Summing up, it can thus be concluded that the (syntactic) Swabian 1SgNom pronoun has two corresponding lexical p-forms, the full form [i:] and the clitic [ə], whose realization depends on their respective status in information structure. A third (optional) empty form of the pronoun, the subject drop, is derived postlexically from the clitic version and is determined by the following phonological conditions: a) the clitic has to be part of a cluster and b) the deletion has to preserve a valid syllable structure.

While these rules and constraints concerning the clitic subject drop clearly apply postlexically (in the sense of Section 4.1.2), the phonological processes involved in the prosodic grouping of the full form [i:] and the clitic form [ə] with respect to the previous element need further clarification – in particular with respect to the question as to which prosodic information is stored in the lexical entry and which is derived postlexically. This will be the topic of the next two sections.

¹⁴ O stands for ‘onset’, N for ‘nucleus’ and C for ‘coda’; the ‘rhyme’ is deleted for reasons of space and simplification.

¹⁵ Note that the coda of the host [n] is ambisyllabic in that it also fills the onset position of the following syllable. However, this consonant is not left astray, because it remains part of the previous syllable’s coda as well.

4.2.2 The prosodic status of 1SgNom pronouns

In order to evaluate the differences between the two p-form entries [ə] and [i:], some basic assumptions about the clitic status of the Swabian pronoun have to be made. As discussed in Section 4.1.1, this thesis follows Anderson (2005)’s threefold distinction of clitics, where a clitic is either syntactically idiosyncratic, prosodically deficient, or both. In this system, the Swabian 1SgNom pronoun clitic would belong to the second group, as its linear position does not vary in comparison to the distribution of the corresponding full form.

Clitics of this group are prosodically deficient, i.e., they need to lean on a prosodic word which is defined as “minimally a stressed foot [...] and maximally a single lexical word combined with any associated unstressed function words” (Wheeldon 2000, 254). Under these assumptions, the Swabian 1SgNom pronoun clitic cannot form a prosodic word in contrast to the corresponding full form which is stressed and can thus form a foot and consequently a prosodic word. Following, among others, Booij (1988), it is assumed that the prosodic structure up to the level of the prosodic word is assigned in the lexicon. As a consequence, this information about prosodic structure must be available as part of the lexical entry, i.e., it needs to be indicated that the strong form [i:] forms a prosodic word and that the clitic [ə] does not do so. The question remains as to how the clitic should be grouped together with the preceding host and if this information is stored lexically as well – or if it is derived postlexically. This is the topic of the following two subsections.

4.2.2.1 The prosodic grouping of ‘function words’ – in general

Selkirk (1995) proposes several possibilities for the grouping of ‘function words’ with their respective hosts ((48), adjusted to show enclitics), among them several examples which violate the *strict layer hypothesis* (Selkirk 1984, 26).

- (48)
- a. prosodic word $((\text{host})_\omega (\text{clitic})_\omega)_\phi$
 - b. free clitic $((\text{host})_\omega \text{clitic})_\phi$
 - c. internal clitic $((\text{host clitic})_\omega)_\phi$
 - d. affixal clitic $(((\text{host})_\omega \text{clitic})_\omega)_\phi$

Several researchers have analysed pronoun clitics in Germanic languages.¹⁶ German personal pronouns have been analysed by Hall (1999) who defines the German prosodic word as minimally bimoraic (Hall 1999, 106), thus excluding all weak object pronouns ending in monomoraic short vowels or syllabified consonants ([sə, ŋ]). Furthermore, a prosodic word has to contain at least one full vowel (Hall 1999, 114), ruling out ə-consonant combinations (e.g., [əs]). A positive rule to identify

¹⁶Gerlach and Grijzenhout (2000) analyse Dutch pronoun clitics as free clitics. Selkirk (1995) assigns affixal clitic status or prosodic word status to English object pronouns depending on their position in the clause. See also Spencer and Luís (2012, 103) for other examples.

prosodic word status is syllabification in cases where the final consonant of the host resyllabifies to the onset of the following vowel-initial clitic.

Concerning the German third person singular feminine nominative pronoun, Hall distinguishes three forms: one strong form [zi:] and two weak forms [zɪ] and [zə], where [zɪ] occurs in normal conversational German and [zə] is possible in more casual pronunciation (Hall 1999, 103). The two weak forms are crucial for his differentiation between two possible integrations of clitics with the preceding host. Based on the *lax vowel constraint* which prohibits short lax full non-low vowels from appearing at the end of a prosodic word, Hall derives a difference in prosodic grouping for the three forms.¹⁷ He generalizes that clitics form a prosodic word together with the host, because the host's final consonant resyllabifies to the onset of a following vowel-initial clitic, as it is the case in combinations like *geht er* ('goes-he'): [ge:.tə]. However, the assumption that the clitic forms a prosodic word with the host is problematic with the weak pronoun form [zɪ], as it ends in a lax full vowel and consequently cannot be positioned at the end of a prosodic word. In order to prevent a violation of the *lax vowel constraint*, Hall thus concludes that this specific clitic must be integrated at the level of the phonological phrase.

In addition to the two weak forms, the third option is the focussed pronoun ([zi:]) which Hall suspects to be an independent prosodic word (Hall 1999, 128) – a suspicion, confirmed (among others) by Wheeldon and Lahiri (1997, 358) who state that “under focus, function words are always phonological words”.¹⁸ As a consequence, there are three types of prosodic phrasing proposed by Hall (1999) for the third person singular feminine pronoun *sie* ('she'), shown in Figure 4.5 in combination with the verb *kann* ('can').

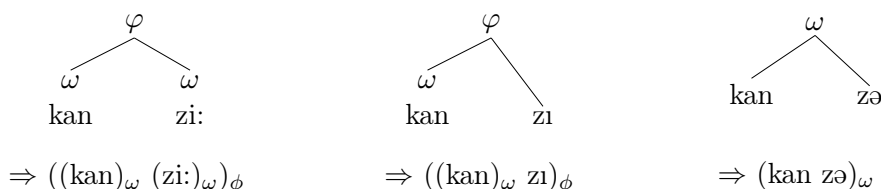


Figure 4.5: Prosodic grouping of 3SgFem pronouns (Hall 1999).

In Selkirk (1995)'s terms, [zi:] is a prosodic word, [zɪ] is a free clitic and [zə] is an internal clitic. The following section discusses if these assumptions for standard German can be applied to the Swabian data as well.

¹⁷Note that Hall excludes ə from the group of ‘short full lax non-low vowels’ and refers to [ɪ ʏ ε œ ʊ ɔ].

¹⁸This is also the reason why there needs to be a distinction between function word and clitic: A function word often is a clitic, but does not have to be.

4.2.2.2 The prosodic grouping of Swabian 1SgNom pronouns

Standard German does not differentiate between a weak and a strong first person singular nominative; the 1SgNom is context-independently always pronounced as [iç]. However, as has been shown in Section 4.2, the Swabian dialect distinguishes two forms: the strong form [i:] and the weak form [ə]. Following the above statements on pronoun clitics and prosodic words, prosodic word status is assumed to be identifiable by domain-specific processes like syllabification. This approach follows (among others) Wheeldon and Lahiri (1997)'s assumption that independent prosodic word status for the full pronoun is indicated by an absent syllabification to the previous word and by a glottal stop before the stressed vowel in context ([?i:]).¹⁹

One could argue that Swabian has at least two weak forms as well. Depending on the speech rate, the pronoun [i:] in the following example ((49)) can be either syllabified to the host or not.²⁰

- (49) den fiʃ wil i: nɛt
 the.DAT fish want.1SG.PRS 1SG.NOM not
 "I don't want that fish."

careful speech: (wil)_ω (?i:)_ω
 fast speech: (wil=1)_ω

The resulting conclusion is that depending on the speech rate, the nature of the pronoun and, more importantly, the prosodic phrasing of a whole sentence can change. However, this does not necessarily imply several different pronoun forms in the lexicon; instead, the respective changes can be derived *postlexically* if triggered by an external factor like speech tempo. Kleinhenz (1998, 169ff) argues that with a variation in speech tempo phonological rules tied to the domain of the prosodic word (e.g., syllabification) can apply to higher units; or, from a different perspective: The prosodic word is 'stretched'. It can thus be concluded that there are not two different clitic forms for the Swabian 1SgNom pronoun (and possibly not two weak forms for the Standard German 3SgFNom pronoun as well), but that the perceived difference in the weak forms result from external factors, e.g., speech rate. To reduce possible confusion, the Swabian clitics are only discussed in an assumed careful speech rate where the fully stressed pronoun in (49) will form an individual prosodic word: (wil)_ω (?i:)_ω.

Since the unstressed clitic pronoun [ə] is subject to syllabification, a process assumed to be constrained by the domain of the prosodic word, there are two possibilities left for the analysis of the host+clitic combination: It has either an internal clitic structure (host clitic)_ω, or it is analysed as an affixal clitic ((host)_ω clitic)_ω. In

¹⁹In German, stressed vowel-initial syllables begin with a glottal stop.

²⁰There is one exception: If the personal pronoun carries contrastive focus, it will always be pronounced as an individual prosodic word. In contrast, the weak pronoun [ə] will never be a prosodic word.

the following, these two possibilities are discussed by looking at an optional Swabian postlexical phonological process called *n-insertion*.²¹

Swabian *n-insertion* optionally applies if a host ending in a vowel is followed by a vowel-initial enclitic. In that case, the speaker can optionally insert a linking *n* to avoid a vowel hiatus ((50a)). This process is ungrammatical between two prosodic words ((50b)).

- (50) a. va:f du: vo:=(**n-**)ə dɛs hap
 know.2SG.PRS 2SG.NOM where=(n-)1SG.NOM this have.1SG.PRS
 ‘Do you know where I’ve got this?’
- b. * vo:)ω **n-** ω(ʔe:fa: vo:nt
 where n-Eva.3SG.F.NOM live.3SG.PRS
 ‘... where Eva lives.’

From these examples it can be concluded that *n-insertion* in Swabian takes place within the prosodic word, thus explaining the difference between (50a) where the *n* is applied inside the prosodic word, and (50b), where *n* is placed between two prosodic words.

Furthermore, *n-insertion* only applies if the prosodic word is nested; i.e., to the left of the insertion, there has to be a right prosodic boundary: ((host)_ω n-clitic)_ω. If the clitic was integrated into the same prosodic word with its host it would be difficult to explain the limitation of the *n-insertion* to this specific place, in that the *n-insertion* rule would not be able to distinguish the space between host and clitic from a host-affix combination or even a stem-internal vowel hiatus. By assuming a nested prosodic word, this limitation of *n-insertion* to the position between host and clitic can be explained. Further evidence for this analysis comes from (51).

- (51) * vo:=ə=**n-ə**m k^hɔlfɐ han
 where=1SG.NOM=n-3SG.M.DAT help.PRF have.1SG.PRS
 ‘... where I helped him.’

Example (51) shows that *n-insertion* is ungrammatical between two adjacent clitics, which could not be explained if the clitics were both included into one prosodic word with the host ((host clitic clitic)_ω). The same is true for the case where each clitic is assumed to project a further layer of a nested prosodic word: (((host)_ω clitic)_ωclitic)_ω. In this case, the *n-insertion* would be valid as well, as it would find a

²¹For a first overview, see Haag-Merz (1996, 97, and references therein). The following assumptions have been derived and confirmed through judgement testing of six speakers who grew up near the towns of Göppingen and Tübingen. Speakers were asked to indicate their preferred choice(s) from a list of several constructions. Note that the speakers of Swabian from the city of Stuttgart do not seem to use *n-insertion*, while *n-insertion* (with slightly different constraints) can be found in Bernese Swiss German Penner (1991), but a geographic distribution of the phenomenon awaits further research.

dependent on the corresponding information structure status with respect to focus;²³ there are no phonological conditions which determine the alternation. Thus, both variants have to be listed in the lexicon. (c.f. Anderson 2005, 27). In contrast, the alternation between the unstressed pronoun subject [ə] and the null subject described in Section 4.2.1 is solely conditioned by postlexical phonological processes. Therefore, the null subject is not listed as a possible variant in the lexicon, but is assumed to be (optionally) derived from the (lexical) [ə] via postlexical rules and constraints. Finally, the prosodic status of the full pronoun on the one hand and the clitic on the other hand were also determined. While the full pronoun is lexically stored as a prosodic word, the clitic does not project any prosodic information above the syllable level. Its prosodic integration into the host takes place postlexically, where the clitic and the host form a nested prosodic word structure. In Table 4.1, the ordered set of postlexical phonological rules and constraints is applied to some of the examples discussed above.

column	1 (ex. (45))	2 (ex. (46))	3 (ex. (47))	4 (ex. (51))
	V-Subj-Obj	V-Subj-Obj-Part	Subj-V-Obj	Adv-Subj
(lexical output)	'hɛn=ə=s	'hap=ə=s=ŋ	'i: 'hɛn=s	'vo:=ə=əm
	↓	↓	↓	↓
<i>pros. phrasing</i>	((hɛn) _ω ə s) _ω	((hap) _ω ə s ŋ) _ω	(i:) _ω ((hɛn) _ω s) _ω	((vo:) _ω ə əm) _ω
<i>syllabification</i>	'hɛn.nəs	'hap.pəs.ŋ	i:.'hɛns	'vo:ə.əm
<i>cluster</i>	✓	✓	—	✓
(<i>subj. deletion – valid σ-structure</i>)	—	'hap.∅.ŋ	—	'vo:∅.əm
(<i>n-insertion</i>)	—	—	—	'vo:=n-əm

Table 4.1: Postlexical phonological rules/constraints and Swabian clitics.

In a first step, the *prosodic phrasing* of each construction is determined on the basis of the lexical prosodic status. After *prosodic phrasing* has applied, a *syllabification rule*,²⁴ groups the segments into valid syllable structures, resyllabifying the final consonant of the host into the onset of the following clitic in columns 1 and 2, and drawing the unstressed [s] into the coda of the verb's syllable in column 3. A further constraint checks if a clitic cluster is present. It is only then that the optional *deletion* (indicated by round brackets (...)) of the subject clitic follows under the condition that each *remaining segment is still part of a valid syllable structure* which is the reason why subject deletion only applies to the second and the fourth, but not the first column. Following this process is the *n-insertion* which only applies to the construction in column 4, as the others don't have a vowel hiatus.

²³Although this dissertation explicitly excludes information structure related prosodic phenomena, an exception is made here in order to motivate the Swabian data analysis. To fully analyse the concept 'information structure in relation to prosody' would go far beyond the scope of this dissertation.

²⁴The following annotation is used to indicate ambisyllabicity: x.x.

Table 4.1 also shows how important the ordering of the postlexical phonological rules and constraints is for the derivation of the correct form. If, for example, the syllabification rule and the deletion of the subject clitic would be reversed, then there would be no reason not to apply subject deletion to the first column, as the resulting stray consonant could be resyllabified into the coda of the host. A consequence of the early application of syllabification is also the constraint that *n-insertion* has to apply *after* subject deletion; otherwise, [n] would have been added to the onset of the subject clitic [ə] in column four and a deletion would not have been possible, because it would result in an invalid syllable structure: a stray [n]. In contrast, if *n-insertion* applies after subject deletion, [n] will be placed in the onset of the object clitic [əm] and the valid syllable structure is retained.

4.2.4 Swabian pronoun alternation: an LFG analysis

The topic of this section is the implementation of the Swabian 1SgNom pronoun in medial clause position into LFG. First, it will be shown how the i(nformation)-structure status of the pronoun determines the corresponding clitic/full p-form in the lexicon. Following this is a section on the transfer of information between lexicon and p-structure. The section concludes with a subsection on Swabian postlexical phonological rules as presented in Section 4.2.3, showing how the preliminary one-to-one transfer from the lexicon to p-structure is transformed towards an individual p-structure representation which qualifies as a partial input to an actual speech signal generation.

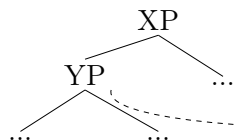
4.2.4.1 Information structure status and lexical selection

King (1997) and Butt and King (1997) introduce information structure as an attribute value matrix to encode discourse-function information, e.g. FOCUS or TOPIC. The resulting projection is related to c-structure by the correspondence function \uparrow_i in that (for example) a c-structure annotation rule as in (54) will *project* the information that all terminal nodes belonging to the YP node in c-structure are part of the i-structure attribute FOCUS.

$$(54) \quad \text{XP} \quad \longrightarrow \quad \text{YP} \quad \dots$$

$$\quad \quad \quad \downarrow_i \in (\uparrow_i \text{ FOCUS})$$

c-structure:



i-structure:



However, with respect to the Swabian data, the annotation is not tied to the c-structure annotation rule for pronouns in general, but to a very specific lexical item: The first person singular nominative (1SgNom) pronoun and its two variants whose phonetic realizations depend only on i-structure status and are thus listed in the lexicon.

CONCEPT	S-FORM	P-FORM
I	i PRON (\uparrow PRED) = 'pro' (\uparrow PRONTYPE) = pers (\uparrow NUM) = sg (\uparrow PERS) = 1 { (FOCUS \uparrow_i)	SEGMENTS /i:/ ----- SEGMENTS /ə/
	\neg (FOCUS \uparrow_i)}	

Table 4.2: Part of the lexical entry for the Swabian 1SgNom pronoun.

Table 4.2 encodes the lexical entry for the first person singular pronoun in Swabian. The S-FORM encodes pronoun-type, number and person of the pronoun which are identical for both variants. Following this is the annotation of a disjunction (LFG encoding: {option1 | option2}) concerning the FOCUS attribute in i-structure (\uparrow_i). In this case, an *existential constraint* in combination with *inside-out functional uncertainty* is used. An *existential constraint* only names a specific attribute (in this case: (FOCUS)) and checks whether the lexical item (the pronoun's s-form *i*) is within the scope of that attribute. The *inside-out functional uncertainty* (indicated by the referral to the mother node (\uparrow_i) following the attribute) allows for the possibility to check for the attribute no matter how deeply the lexical item is embedded in the attribute's structure. Without its application (as in (\uparrow_i FOCUS)), the existential constraint would only check for the attribute FOCUS in the immediate structure enclosing the lexical item - which is (possibly) not the structure that contains the FOCUS attribute, but one of the structures that are nested into it (e.g., the undefined attribute ATTR in Figure 4.6).

The disjunction in the lexical entry encodes two possibilities: Either there is a FOCUS attribute somewhere in the mother structure of the pronoun or there is none, indicated by the negation operator \neg . In the former case, the stressed p-form is chosen (/i:/), in the latter case, the unstressed form /ə/ is realized.

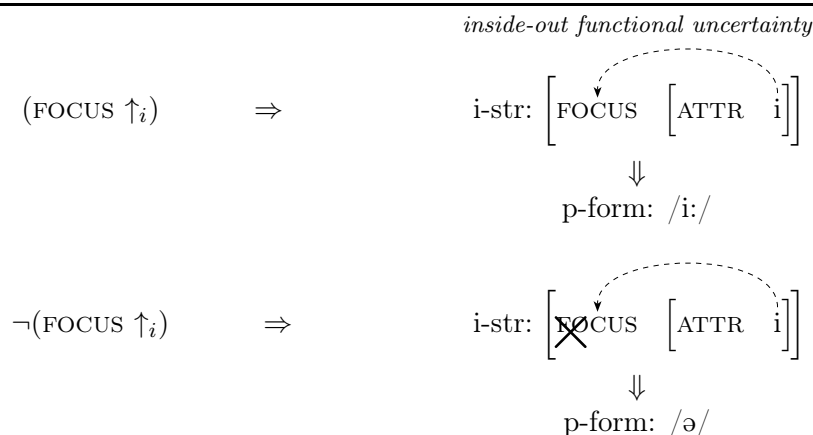


Figure 4.6: Existential constraint and inside-out functional uncertainty.

Figure 4.6 illustrates this selection process. In this i-structure representation²⁵ the FOCUS attribute contains further subdivisions, in this case only one underspecified attribute ATTR whose value is the s-form of the pronoun *i* ('I'). The constraints formulated on the left side check whether the FOCUS attribute is part of the s-form's mother structure in i-structure. In the first part of Figure 4.6, this is the case and as a consequence, the p-form /i:/ is realized; in the second part, the FOCUS attribute is not part of the mother structure. Thus, the corresponding p-form /ə/ is realized. Following this selection process, the respective p-form's phonological information is available for further processing.

4.2.4.2 From lexicon to p-structure: the *transfer of vocabulary*

The following section discusses the transfer of information from the lexicon to p-structure. As mentioned in Section 4.1, the general transfer process between the syntactic and the phonological component is assumed to be twofold: the *transfer of vocabulary* and the *transfer of structure*. While the *transfer of structure* will be in the focus of the following Chapters 5 and 6, the current chapter describes the *transfer of vocabulary*: from lexicon to p-structure. In that context, the term *vocabulary* represents the lexical material associated with each s- and p-form. If the interface transfers information from c-structure to p-structure, the lexicon 'translates' the incoming s(yntactic)-form to the corresponding p-form. The p-form and the associated phonological material are then transferred to the (preliminary) p-diagram, the input to p-structure.

Following the conclusions of Section 4.2.2.2, it is assumed that the pronoun's full form projects an independent prosodic word, while the pronoun's clitic form does not do so. The following table focusses on the phonological part of the lexical entry

²⁵Note that this thesis does not make any claims on how i-structure should be structured. A focus-value following Krifka (2008) and a possible further division of the FOCUS attribute (along the lines of Andréasson (2007)), which is encoded as an underspecified attribute ATTR, is assumed.

of Table 4.2 and extends it according to these findings.

S-FORM	P-FORM
...	
{(FOCUS \uparrow_i)}	P-FORM [i:]
	SEGMENTS /i:/
	METR. FRAME (' σ) $_{\omega}$

\neg (FOCUS \uparrow_i)}	P-FORM [\emptyset]
	SEGMENTS / \emptyset /
	METR. FRAME = σ

Table 4.3: P-form of the Swabian 1SgNom pronoun.

The p-form in Table 4.3 contains two possible variants: focussed [i:] and unfocussed [\emptyset]. The former's metrical frame contains one stressed syllable, indicated by ' σ ', and is interpreted as a prosodic word (also cf., among others, Wheeldon and Lahiri 1997, 358). The latter has an additional annotation next to the syllable: = σ . This annotation²⁶ indicates that the respective p-form is prosodically deficient and has to lean on a host to its left, i.e., it cannot be interpreted as a prosodic word.²⁷

An alternative solution (or a possible extension) would be the subcategorization for prosodic frames as it was proposed by Inkelas (1990) where the clitic subcategorizes for a specific prosodic environment in the lexical entry, e.g. ((...) $_{\omega}$ $_{-}$) $_{\omega}$ for nested prosodic words. This approach evades the necessity for an explicit postlexical rule that determines prosodic phrasing on the basis of the prosodic information projected from the lexical entry. However, phrasing rules will continue to be necessary as the higher prosodic domains are, among other conditions, influenced by syntactic phrasing (e.g., in the case of parentheticals).

Subcategorization frames are especially helpful if there are a number of clitics in the language whose phrasing takes place at the level of different prosodic domains, as these differences cannot be captured by the = σ annotation alone. A drawback of the approach is that the description of clitic clusters requires some extra formal power as these are not stored in the lexicon. For the current phenomenon in Swabian, an implementation of prosodic subcategorization frames is not needed, as there are (to my knowledge) no clitics incorporated into levels above the prosodic word – thus, an integration of prosodic subcategorization frames into the lexical entry is unnecessary. Nevertheless, this possibility should be kept in mind for languages featuring different types of clitics.

²⁶The annotation = imitates the glossing of clitics as proposed by the *Leipzig Glossing Rules*.

²⁷This annotation also explains why the clitic form cannot appear clause-initially: The clitic cannot find a suitable host to its left. However, since this discussion is restricted to the medial clause position, this is not further evaluated here.

The following example will be used for demonstration purposes. It includes a clitic cluster allowing for a subject drop as well as for *n-insertion*.

- (55) vaɪf du: vo:(=(n-)ə)=sə ñ: han
 know.2SG.PRS 2SG.NOM where(=(n-)1SG.NOM)=3SG.F.ACC there(.put.PRF) have.1SG.PRS
 ‘Do you know where I put her?’

The s-forms of example (55) are related to the following p-forms:

P-FORM	[vaɪf]	[du:]	[vo:]	[ə]	[sə]	[ñ:]	[han]
SEGMENTS	/v a ɪ f/	/d u:/	/v o:/	/ə/	/s ə/	/ñ:/	/h a n/
METR. FRAME	(σ) ω	(σ) ω	(σ) ω	= σ	= σ	(σ) ω	(σ) ω

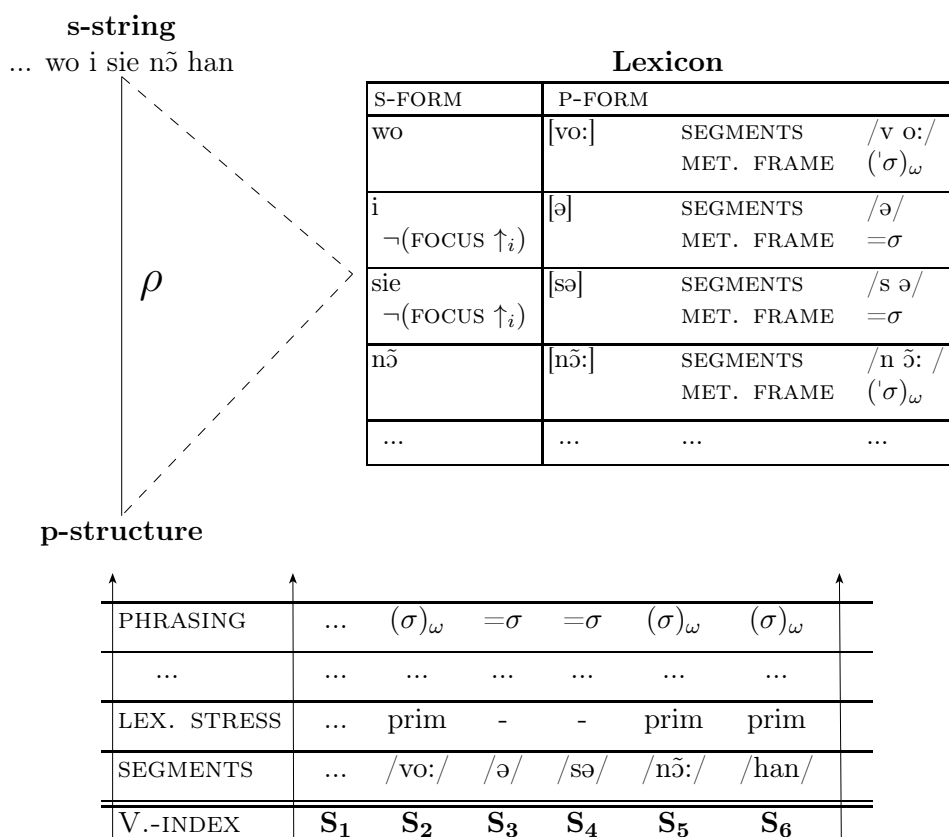
Table 4.4: P-form entries corresponding to example (55).

The segments are distributed among the syllables according to syllabification rules before they are encoded in the p-diagram. Note, however, that this is a simplification at that point. In the lexical entry of the p-form, the metrical structure is separated from the segments. The reason for this is that syllabification may change with a different phonological environment and should thus be treated as a postlexical process. A similar discussion is led by Booij (1988) who establishes that prosodic structure is part of the lexical entry – and syllabification a postlexical process bound to the domain of the prosodic word. This general assumption is confirmed by the postlexical processes described in this chapter as they are set within the domain of the prosodic word and are thus subject to postlexical syllabification. As a consequence, the lexically transferred material should be unstructured, syllablewise. However, for reasons of simplification (and because the postlexical status of syllabification cannot fully exclude a parallel lexical status of syllabification), the lexical material is ‘pre’-syllabified according to the isolated lexical form.

This syllable structure of the isolated lexical entries is then transferred to the p-diagram via the relation ρ which relates the (syntactic) string to p-structure via the lexicon (see Chapter 3, Section 3.5). Thus, in addition to the previously introduced *interpretation* level and *signal* level, a third level can be encoded in the p-diagram: the *lexical level*.²⁸ Figure 4.7 shows the transfer from the lexicon to a preliminary version of p-structure.²⁹

²⁸However, note that the *lexical* and the *signal* level are mutually exclusive, as the former applies only during production and the latter during comprehension.

²⁹Note that the syntactic structure information usually feeds into the preliminary version of the p-diagram as well (see Chapter 3, Section 3.5). However, the transfer of syntactic structure to higher prosodic units is excluded from this chapter as it is not essential for the analysis of the phenomenon discussed here.

Figure 4.7: *Transfer of vocabulary*: metrical, segmental, and prosodic information.

The p-diagram in Figure 4.7 represents the *preliminary* form of p-structure after the relevant lexical information has been transferred. The attribute SEGMENTS encodes the linear order of the s-string translated into syllables according to the metrical frames of the corresponding p-forms. Each of the columns is indexed in order to keep track of the phonological information related to each syllable. This information consists of the attributes like, e.g., LEXICAL STRESS which encodes the lexical stress pattern related with each syllable. In Figure 4.7, all syllables either have primary stress or are unstressed.

The last attribute depicted in Figure 4.7 is PHRASING and refers to prosodic phrasing, in this case only the prosodic words and the clitics, i.e., the lexical prosodic status information of each p-form. Prosodic word status projected from the lexicon is indicated by the notation (...)_ω. Clitics don't have prosodic word status and are thus only projected as syllables³⁰ where the = annotation indicates the direction of cliticisation. The current version of the p-diagram only consists of lexical input

³⁰In the exceptional case of the neutral third person singular pronoun [s], the representation would be monomoraic, i.e., smaller than a syllable. After postlexical syllabification, [s] would then be part of an adjacent syllable structure.

and is thus a (relatively) isomorph representation of the lexicon. It is then the responsibility of the postlexical phonological rules and constraints to transform this preliminary p-diagram into a (partial)³¹ output of p-structure.

4.2.4.3 Postlexical phonology in p-structure

The concept of postlexical phonological rules has been introduced in Section 4.1.2 and a summary of the postlexical rules and constraints relevant for the analysis of the Swabian 1SgNom pronoun was given in Section 4.2.3. This section is concerned with the formalization of these rules and constraints and the related subsequent transformation of the preliminary version of the p-diagram into the final p-diagram which is the output of p-structure. Note, that the preliminary p-diagram, the postlexical phonological rules component and the final p-diagram output are all within the domain of p-structure; thus, the following discussion is no longer concerned with the syntax-prosody interface and how information is transferred, but with the subsequent processing of the transferred information via postlexical phonological rules. For the application of these rules, the formalizations proposed in the phonological literature (represented in (56)) will by and large be applied.

$$(56) \ x \longrightarrow y / V _ C$$

The abstract rule in (56) is interpreted as ‘x becomes y (\longrightarrow) in the context (/) of a preceding vowel (V) and a following consonant (C)’, where $_$ is the placeholder for the element in question. This basic rule can be expanded in numerous ways; a quick overview on possible formal extensions is given in Table 4.5.

³¹Whereby ‘partial’ refers to the fact that a complete p-structure cannot be derived solely from lexical and syntactic information. Other very important factors are, e.g., semantic and pragmatic contributions.

symbol	explanation	symbol	explanation
$\begin{bmatrix} x \\ y \end{bmatrix}$	conjunction: x AND y	$\left\{ \begin{matrix} x \\ y \end{matrix} \right\}$	disjunction: x OR y
[+x]	feature x is present	[-x]	feature x is absent
(x)	x is optional	\emptyset	empty element
σ	syllable	$(\sigma)_{\sigma}$	left and right σ boundary
ω	prosodic word	$(\omega)_{\omega}$	left and right ω boundary
ϕ	phonological phrase	$(\phi)_{\phi}$	left and right ϕ boundary
$x_i / y_i _$	x copies features of y	*xy	constraint: xy is not allowed
...	anything	C'	unsyllabified consonant
α	variable	?	any phoneme
x^*	x appears 0- ∞ times	x+	x appears 1- ∞ times
x^n	x appears n times	?*	any material

Table 4.5: Phonological rule annotations used in this dissertation

While most of the rules in Table 4.5 are standard annotations that can be found in numerous textbooks (see, e.g., Hall (2011), Hayes (2009) and Odden (2005)), the last five annotations are unusual in theoretical phonology. These are regular expressions used (among others) by Beesley and Karttunen (2003) to express conditions for (phonological) rules in their finite state transducers, which are not easily captured in standard phonological rule annotation. As the theoretical approach proposed in this thesis is also considered from the perspective of a prospective computational implementation, these annotations are applied in the following phonological rules as well.³²

A first phonological rule has been expressed in (53) which described the phenomenon of optional *n-insertion* in Swabian. In the following, the set of rules and constraints operating on the part in bold letters in example (57), (repeated from (55)), will be formalized and explained.

(57) vaɪf du: **vo:(=(n-)ə)=sə** n̄ɔ: han
 know.2SG.PRS 2SG.NOM where(=(n-)1SG.NOM)=3SG.F.ACC there(.put.PRF) have.1SG.PRS
 ‘Do you know where I put her?’

Based on the paradigm developed for the 1SgNom clitic pronoun in Sections 4.2.1 and 4.2.2.2, the following phonological rules and constraints apply to **vo:(=(n-)ə)=sə** in the given order.

³²Note that ?* is an annotation frequently used in the following chapters. While ?* in a strict interpretation means ‘anything any times’, it is used in this thesis with the meaning ‘anything, except the material under discussion’. Usually, this specification has to be encoded by adding further filters, but this is neglected for reasons of simplification.

1. **lexical input to p-structure:** $(\omega \text{vo:})_{\omega} =\emptyset =\text{s}\emptyset$

Following the discussion and conclusions in Section 4.2.2, this lexical input consisting of a prosodic word and two clitics is first phrased together into a nested prosodic word.

2. **prosodic phrasing:** $\omega_i = (?+)_\alpha^n \longrightarrow (\omega \ \omega_i \ (?+)_\alpha^n)_{\omega}$

The prosodic phrasing rule in 2. can be read as follows: If there is a prosodic word followed by one or several clitics³³, then phrase the prosodic word together with the clitics into a nested prosodic word construction. The rule under 2. is partly iterative: $(?+)_\alpha^n$ indicates that there can, in principle, occur n clitics, where the variable α represents the copy index of the currently processed clitic, ensuring the correct linear order of the clitics on the right hand side.

3. **subject-deletion:** $(\emptyset \longrightarrow \emptyset) /)_{\omega} (\sigma _)_{\sigma} \sigma +)_{\omega}$

The above rule (3.) can be read as ‘the syllable \emptyset is optionally deleted if it is directly following a prosodic word boundary in a nested prosodic word construction and if one (or more: $\sigma+$) syllables follow within the same prosodic word domain’. Thus, the *subject-deletion* in 3. also includes the constraint that a subject has to be part of a clitic cluster of at least two syllables (where the syllable borders are indicated by [.]). There are two possible output forms: **[vo:.ə.sə]** or **[vo:..sə]**. In the former, the subject clitic is retained while in the latter, it was deleted.

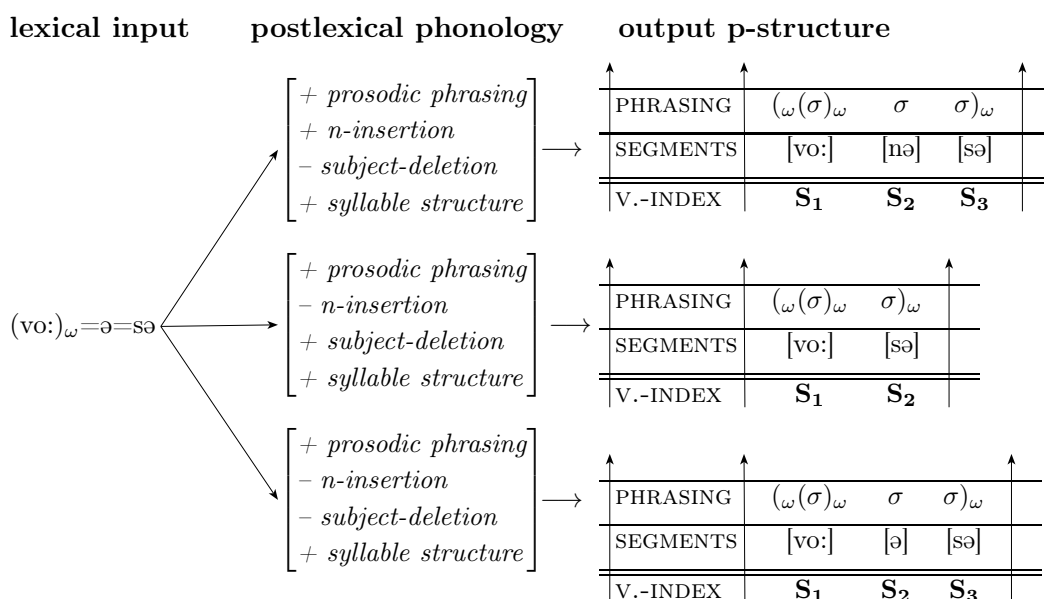
4. **n-insertion:** $(\emptyset \rightarrow [n]) / (\omega (\omega \ ?^* \text{V})_{\omega} (\sigma _ \text{V} \dots)_{\omega}$
where $?^*$ refers to anything except a prosodic word boundary.

The *n-insertion* rule given in 4. places the n into the onset of the following syllable. *N-insertion* is an optional process and consequently, the output can be twofold: **[vo:.n-ə.sə]** or **[vo:.ə.sə]**. In the case of a construction where a subject drop has occurred, *n-insertion* can only apply if the following clitic starts with a vowel as well.³⁴ In the current example, this option is not given.

Summing up, the application of the rules 2–4 to the input given in 1 leads to three possible constructions: **[vo:.n-ə.sə]**, **[vo:.ə.sə]** or **[vo:..sə]**. Figure 4.8 shows how (a compressed) part of the preliminary p-diagram in Figure 4.7 is transformed via the application of postlexical phonological rules into three corresponding valid p-diagrams. The resulting p-diagrams are the output of p-structure and serve as (one of several) inputs for the production of a speech signal.

³³Note that the clitics do not necessarily form full syllables, as it is the case with the clitic $=\text{s}=\text{}$ (‘it/that/the’). Therefore, the $=(?+)$ annotation is used, which can be read as: A clitic consisting of any combination of phonemes, but at least one phoneme.

³⁴Such a case would be the construction that contains a third person singular masculine dative clitic $=\text{ə}m$: **[vo:..sə.n-əəm]** (‘where-I-him’), see also example (52).

Figure 4.8: Postlexical phonological rules applied to $[vo: = \text{ə} = s\text{ə}]$.

4.3 Optimality Theory

The postlexical phonological rule theory introduced in Section 4.1.2 (and later applied in Section 4.2.4.3) implies the use of phonological rewrite rules. However, there is another possibility which is much more commonly applied in current phonological literature: Optimality Theory (OT) (Prince and Smolensky 2004, McCarthy and Prince 1993). The following sections give a quick introduction into the theory behind OT, OT approaches proposed in LFG, and, more importantly, discuss why OT is not applied in this thesis.

OT does not apply phonological rules but determines a (phonological) surface form on the basis of different constraints ranked in an optimal order. OT is based on the assumption that all languages of the world share basic, but conflicting, tendencies which can be expressed by constraints (CON). Although constraints are considered to be universal (a view supported by typological findings), a constraint may be more or less significant in a specific language: The constraints are ranked differently with respect to each other across languages and thus form language-specific hierarchies. As a consequence, the same input leads to different optimal expressions in different languages.

OT distinguishes between two types of constraints within CON: *markedness constraints* and *faithfulness constraints*. *Markedness constraints* are the constraints that fulfill linguistic requirements, e.g., the constraint that consonant clusters are suboptimal. They apply to the surface structure of the output. The notion of *faithfulness* on the other hand requires the output to be as similar as possible to the input, i.e.,

this class of constraints is concerned with the relationship between input and output. As *markedness constraints* support changes to a given input to fulfill linguistic requirements, they are often in tension with *faithfulness constraints*, as the latter attempt to preserve the input in its original form.

An important factor of OT is that the constraints are violable. An output form may have violated several constraints – however, as long as these violations are on a lower level of the constraint hierarchy compared to the other available options and related violations on a higher level of the constraint hierarchy, these violations are tolerated.

The basic idea behind this concept is that there is a possibly infinite number of output forms for a given input which are generated by the function GEN (‘Generator’). These output forms are then evaluated (EVAL) according to the constraint hierarchy of a specific language. The output form which violates the least and lowest constraints is the *optimal candidate*. An abstract example is given in Table 4.6, where the higher ranked constraint A is placed to the left of the lower ranked constraint B ($A \gg B$). The (in principle) infinite number of candidates generated in GEN are represented by only two candidates in this table. A constraint violation is indicated by *, a violation which causes the candidate to be ruled out is indicated by *!, and a constraint fulfillment is indicated by an empty cell.

Input	A	B
candidate 1	*!	
☞ candidate 2		*

Table 4.6: An abstract example for Optimality Theory.

Although candidate 2 violates constraint B, it is marked as optimal (indicated by ☞) because candidate 1 violates the higher ranked constraint A which is fulfilled by candidate 2. In more general terms that means that OT does not take an input and transforms it according to phonological conditions into an output, but instead takes an input paired with *all possible* (and possibly infinite) outputs and then analyses these outputs according to a hierarchy of conflicting (phonological) constraints. The determination of the correct output form is thus not part of the generation (GEN) (as it is the case with postlexical phonological rules), but is postponed until evaluation (EVAL), where the constraint hierarchy CON is applied to reduce the output candidates (over)generated in GEN to one *optimal* candidate (which can, in principle, be ungrammatical).

This separates the OT approach from, e.g., traditional LFG grammars or the theory of phonological rules, where unviolable constraints (the grammar) determine the rejection or the acceptance of a given input in what can be called GEN. As a consequence, an analysis might not provide an output: Parsing an ungrammatical string will result in a failure, which, on the other hand also means that an unviolable grammar can deal with ineffable structures, in contrast to OT. Furthermore, parsing

with an inviolable grammar can also return more than one option in cases of ambiguity. OT, on the contrary, always provides an ‘optimal’ candidate, even though this output might be ungrammatical or only one of several.

In addition the issues of ineffability and ambiguity, the main problem when working with OT is that of computational intractability, i.e., the creation of a potentially infinite set of candidates. Numerous papers discuss this issue from different perspectives, e.g., Idsardi (2006) and references therein. Others object to this claim, stating that OT is computationally tractable if certain assumptions are made about the nature of the constraints, i.e., if they are ‘fixed’ and not variable (Heinz et al. 2009).³⁵

In the more theoretical approaches to OT, considerations of computational intractability are mostly ‘dismissed’, arguing that it is “not incumbent upon a grammar to compute” (Prince and Smolensky 2004, 233) which refers to the position that researchers should be free to hypothesize without being restricted by meta-constraints like computational tractability or human memory. These meta-constraints are part of the traditional notion of ‘performance’ and treated separately from what has been named ‘competence’ (Chomsky 2006). And indeed, it is this distinction that is used to argue against the point of computational intractability by the defenders of Optimality Theory.

Chomsky (2006, 102) defines *competence* as the “system of rules that determine both the phonetic shape of the sentence and its intrinsic semantic content” while *performance* involves further facts, e.g. memory, beliefs concerning the speaker and the situation, or mechanical aspects in the production of sounds. As McCarthy (2007) notes, Chomsky states that the rules that relate sound and meaning in a certain order within the *competence* model should not be assumed to be the description of the successive processes of a corresponding *performance* model, i.e., “the grammatical rules [...] do not constitute a model for the production of sentences” (Chomsky 2006, 104). McCarthy (2007) classifies computational tractability as being part of the *performance* and Optimality Theory as being part of *competence*, thus emphasizing the hypothetical nature of OT.

However, Chomsky also remarks that “any such model [=performance] must incorporate the system of grammatical rules [=competence]” (Chomsky 2006, 104, remarks in square brackets added by me), i.e., if a *performance* model is created, then the grammatical rules derived in a *competence* model must be applicable. This is exactly the point where the papers concerned with OT implementations are troubled, as they are bound to create at least a finite set of output candidates to be evaluated by the constraints in EVAL; otherwise, the computation would be impossible. And even though the basic idea that there are no “grounds [...] for assuming that computational complexity considerations, applied directly to grammatical formalism, will be informative” (Prince and Smolensky 2004, 233) is true in the sense that hypotheses on the *competence* model should not be completely determined by

³⁵See Sections 4.3.1 and 4.3.2 for some of the restrictions proposed in the context of OT-LFG.

computational (or, e.g., by cognitive, among others) considerations, the limits of these fields should nevertheless give a clue as to where the rough boundaries of a *competence* model should be.

Leaving this discussion aside, the following section returns to the framework of LFG and shows how OT can be brought together with LFG (syntax) by restricting the candidate set to a (computationally) manageable size.

4.3.1 Optimality Theory in LFG

There are some approaches to bring together OT with LFG syntax. This section provides a short overview on optimal theoretic LFG (OT-LFG) and discusses the (partial) solutions to some of the issues introduced in the previous section.

Coming from the perspective of production (from meaning to form), the INPUT is assumed to be a feature structure representing the main grammatical information without the language-specific surface forms (Bresnan 2000, Sells 2001). On the basis of that input a set of CANDIDATES is generated which consists of c-structure/f-structure pairs (and possibly more structures). The candidate set is generated with a highly unrestricted grammar consisting only of inviolable constraints, e.g., the basic principles of *coherence* and *completeness* which are underlying to every LFG grammar;³⁶ these constraints are ranked highest and their violation will rule out any candidate. Note that this mechanism already restricts the set of candidates to a finite amount.

The candidate set is subsequently evaluated in the EVAL section which contains the violable constraints. For example, faithfulness constraints ensure that the feature structures of the candidate set express the content of the INPUT. Further groups of constraints include categorization constraints that regulate the correspondence between c-structure and the terminal string, and correspondence constraints between a- and f-structure and c- and f-structure.³⁷ The resulting optimal candidate is a pair of c-/f-structure.

Although the candidate set is already constrained by the underlying grammar, the problem of complexity remains when compared to traditional grammars; it is inherent to any OT approach:

The apparent reason why candidate analyses of considerable size have to be constructed prior to optimization (which will typically rule out all but one analyses) is the following: unlike a system with hard constraints, in OT one cannot discard an analysis on the basis of a local constraint violation, since the analysis may still be the best of all possible ones due to more highly ranked constraints.

(Kuhn 2001, 355)

³⁶See, e.g., Dalrymple (2001) for an overview on these principles.

³⁷As this thesis does not include OT in the analysis, the constraints associated with OT-LFG are not explained in detail. See Bresnan (2000) for a thorough discussion.

In addition to the grammar consisting of unviolable constraints, further mechanisms have been developed in order to restrict the set of candidates to a manageable (machine-processable) size. Some of the mechanisms are applied to the INPUT to reduce the number of *unfaithful* candidates in the output: Lee (2001), for example, adds context information to the underspecified feature structure; Kuhn (2001) also proposes an enrichment of the INPUT by adding f-structure features.

In addition to these restriction mechanisms applied to the INPUT, each generated f-structure is subject to the *subsumption requirement* during the process of GEN, which states that each f-structure is subsumed by the input, i.e., it can contain the same or more, but not conflicting information (Kuhn 2001). Yet another possibility to reduce the candidate set would be to profit from locally restricted OT competition in that the input can be split to reduce the number of combinatorial possibilities in the generated candidate set. The resulting optimal candidates can then be combined again to form the final optimal candidate.³⁸ However, even with the application of these methods, the GEN section will create a vast amount of possibilities.

Other issues are ambiguity/optionality³⁹ and ineffability, i.e., the inability of the OT approach to recognize ungrammatical input, because an ‘optimal’ candidate is always chosen. In contrast to traditional OT approaches, ineffability of a given string can be determined via bidirectional optimization in OT-LFG, considering both directions, production and comprehension. The OT-LFG approach described so far has discussed the OT application from the viewpoint of production. However, there are also discussions from the viewpoint of comprehension using the same underlying LFG grammar consisting of inviolable constraints. In this case, the string is the input. Following Kuhn (2001), the string is first parsed into f-structures. From these f-structures, the underlying input is extracted. At this point, the application ‘changes the direction’ and proceeds as if coming from the production side. Thus, from the extracted input, a set of candidates consisting of c-/f-structure pairs is generated and evaluated, determining an optimal candidate. The c-structure of the optimal candidate is then matched against the input string. If the string is grammatical, the linear precedence of string and c-structure should match; otherwise it can be assumed that the string was ungrammatical. Thus, the application of bidirectional OT ensures that the entered string is actually grammatical. On the contrary, if the string is only tested from the comprehension side, it will always yield an optimal candidate, because a winner must always be provided, although the string itself may be ungrammatical.

Given these considerations for OT-LFG, the following section addresses a possible application of OT to the prosody-syntax interface in LFG as it is proposed in this thesis.

³⁸See Kuhn (2001) for a more detailed discussion of this method with respect to matrix and subordinate clauses.

³⁹See Asudeh (2001) and references therein for a detailed discussion of *partial ordering* and *stochastic optimality theory* to resolve ambiguity/optionality issues.

4.3.2 OT applied to the LFG prosody-syntax interface

When applying OT to the prosody-syntax interface in LFG, the phonological component has to be taken into consideration as well. Formally there are two choices: a) One complete OT system is assumed for both, the syntactic and the phonological component, and b) two systems are assumed, one for the syntactic, and one for the phonological component.

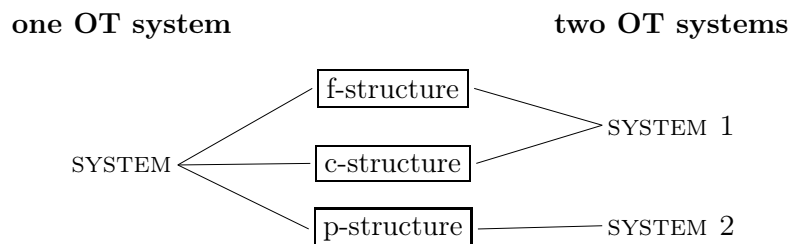


Figure 4.9: Two possible integrations of OT at the prosody-syntax interface in LFG.

Both options assume the same input: an underspecified feature structure as described in the previous section. For the first option, additional structures have to be added to the pair of f- and c-structures generated by the underlying grammar, e.g., p-structure. The underlying grammar which generates the candidates has to be extended to include phonological and prosodic information as well. However, in contrast to the underlying syntactic grammar, the generation of phonological and prosodic information needs much more flexibility. For example, for the generation of prosodic structure only few inviolable constraints can be formulated: E.g., it can probably be assumed that each sentence is contained in an intonational phrase. However, constraints on prosodic units below the level of the intonational phrase cannot be determined as easily, as they are mostly non-isomorphic with syntactic phrasing and thus must allow for a certain flexibility depending on the phonological context. As a consequence, faithfulness constraints that realize the input as closely to the output as possible are not necessarily inviolable and can thus not be used to constrain the candidate set in GEN, but have to be part of EVAL, as violable constraints.

This is also true for the lower prosodic levels and the generation of segmental combinations. Taking the linear order of the c-structure as INPUT, a potentially infinite number of candidates has to be generated. In order to implement this relationship between input and infinite output candidates, the infiniteness has to be restricted with rules that, e.g., relate each consonant of the input to a consonant of the output; otherwise, this process would be impossible. Problematic in this context are processes like Swabian *n-insertion*, where a consonant is added postlexically into a specific position to avoid a vowel hiatus (Section 4.2.2.2). In order to end up with final candidates that also include phenomena like *n-insertion*, all consonants or none would have to be allowed to enter the onset of any vowel-initial syllable.

Karttunen (1998) implements a computational OT system for a CVCV syllab-

ification via the XEROX finite state calculus, which provides interesting insights into the amount of data that needs to be stored. In short, Karttunen first creates a candidate set via the application of a highly unrestricted finite state network. These candidates are then subjected to a constraint section (e.g., ‘*a nucleus position must be filled*’) to filter out the optimal candidate. The constraint hierarchy is modelled with the help of an operation named *priority union* which allows for a certain flexibility: If a candidate is ruled out, because of a constraint violation, it may still be in the output (i.e., it may still be the optimal candidate) if, and only if, there is not at least one other candidate where the constraint violation did not occur. Otherwise, the candidate is ruled out and the unviolating candidate is passed on.

While Karttunen shows that OT constraint hierarchies can, in principle, be modelled via finite state methods, the paper also contains two important conclusions for the current discussion: First it has to be noted that the set of constraints applied has to be ordered in a cascade, i.e., the constraints are not applied in parallel, but if a candidate ‘passes’ one constraint, it is then subject to the following one. Second, while the finite-state calculus proves to be an immensely powerful mechanism, Karttunen gives absolute numbers to two candidate sets: For the output of *a*, the candidate set contains 14 members; for the output of *abracadabra*, however, this number rises close to 1.7 Million candidates – and note that this is a network describing a simple process of CVCV syllabification. What will the numbers look like with more options added to the network?

As can be seen from this example and the previous discussion on prosodic phrasing, constraining the candidate set from the phonological perspective is much more difficult than from a syntactic perspective. If assuming one OT system for both, the syntactic and phonological component, the consequence is an exponentially growing, possibly infinite candidate set.

This problem can be reduced if two separate OT systems are assumed. In this approach, the syntactic structures are generated and an optimal candidate is chosen. The resulting optimal c-structure is taken to be the input to the second OT system, as the linear order of the terminal nodes provide an input string and the tree structure provides some initial phrasing for larger prosodic components.⁴⁰ This approach does not only allow for a reduction of the candidate set by isolating two (and possibly more) components in separate OT systems; it also allows for a modular approach to the problem that was not given by the one-system-for-all approach, where syntactic and phonological constraints can be mixed in principle. But the problem of intractability remains: From a computational perspective, several, to some extent unintuitive restriction mechanism have to apply to reduce the candidate sets created in GEN to a finite number and considerable memory space would have been used to create, store and evaluate sets of candidates.

Nevertheless, the attractiveness of OT’s violable constraints cannot be denied

⁴⁰This approach ‘ignores’ the s-string as a possible input, as the s-string would provide the linear order and the morphosyntactic items, but cannot reflect syntactic phrasing.

as traditional grammars are subject to all-or-nothing analyses, i.e., they consist of inviolable constraints which either reject a candidate or allow for several equally ranked candidates in a case of optionality.

However, there have been some successful approaches to soften up the system of inviolable constraints of traditional LFG grammars. Frank et al. (1998), for example, apply OT-like constraints to a LFG grammar, not in order to find the most optimal candidate, but to indicate preferences between several grammatical analyses. This possibility to add OT-like constraints allows for a more flexible and ranked treatment of candidates (see Chapter 3, Section 3.7.5). A similar method has been found for finite state methods, where weighted transducers allow for preferences in the output (see Mohri (2004) for an overview).

With the notion of ranking combined with traditional grammars, there is, from the perspective of this thesis, no reason to prefer OT to phonological rewrite-rules as introduced earlier in the present chapter: Rewrite rules do not (over)generate a set of candidates and subject it to a set of constraints, but apply directly to the input, generating step by step all possible surface forms. Furthermore, constraints in form of filters can be interspersed with rewrite rules to disallow possible overgenerations.

Returning to the Swabian data in Section 4.2.4.3, the generation of the possible surface forms for the given input is subject to a handful of rewrite rules and constraints. In contrast, the formal power needed to generate the set of candidates according to classic OT assumptions cannot be determined off-hand – this is only possible if the candidates generated in GEN are restricted by further mechanisms and inviolable constraints before they are further evaluated by a (violable) constraint hierarchy: Thus, OT overgenerates candidates before restricting them in contrast to just generating candidates, as it is the implicit idea of phonological rewrite rules. After considering these aspects of computational implementation it can thus be concluded that phonological rewrite rules provide an easier and much more efficient possibility of determining the possible output forms of a given (phonological) input. Furthermore, with the application of OT-like constraints to traditional systems, the obstacle of the inviolable grammar is resolvable for those cases where it is needed.

4.4 Summary and conclusion

This chapter showed how the different components of grammar interact in the analysis of Swabian 1SgNom pronouns in medial clause position with a focus on the interaction between lexicon and p-structure on the one hand and the application of postlexical phonological rules within p-structure on the other hand.

The lexical entry of the Swabian 1SgNom pronoun distinguishes between two p-forms: a fully stressed form [i:] and an unstressed clitic form [ə], which are both part of the lexicon and project different prosodic structures. The projected information is stored in a preliminary p-diagram which is then subjected to postlexical phonological rules. Among these are rules that rephrase the prosodic information

projected from the lexicon, insert [n] to avoid a vowel hiatus between two items, or derive the third variant of the 1SgNom clitic pronoun: the optional subject drop which applies if certain phonological conditions are met and which is thus not part of the lexicon, but a postlexical variant of the clitic pronoun. These postlexical phonological rules are applied to the preliminary p-diagram and generate the final version(s) of the p-diagram which consist of the lexical, (syntactic) and postlexical phonological information and are the output of p-structure.

While this chapter has been concerned with the *transfer of vocabulary*, i.e., the interaction between lexicon and p-structure, the following chapter will focus on the *transfer of structure*, showing how syntactic structure can be transferred to p-structure to give a first indication of the prosodic phrasing of a sentence in production. Furthermore, the following chapters also discuss an important question underlying to the differentiation between syntactic and prosodic/phonological structure: Is the linear order of the items given in c-structure congruent to the linear order given in the phonological structure? The following chapters show that this is mostly the case – but that there are exceptions.

Chapter 5

Endoclisisis at the interface: Degema

5.1 Introduction

The previous chapter introduced the *transfer of vocabulary* from the perspective of production, i.e., the transfer of lexical phonological information (including prosodic information on the level of the prosodic word and below) from the c-structure/the syntactic string to p-structure. A more fine-grained view of p-structure was provided, where a preliminary p-diagram was generated via a transfer of lexical phonological information. This preliminary p-diagram was then subjected to a set of postlexical phonological rules and constraints and transformed into the final p-diagram, the output of p-structure.

This chapter builds on these findings and introduces the *transfer of structure* from syntax to prosody, i.e., the transfer of syntactic structure information to corresponding prosodic domains above the prosodic word: the phonological phrase (φ) and the intonational phrase (ι). As with the previously introduced *transfer of vocabulary*, the *transfer of structure*'s destination is the preliminary p-diagram which serves as the input to p-structure, and can thus be seen as an isomorphic 'translation' of syntactic structure into prosodic structure. It is within the domain of p-structure that this preliminary 'prosodic' phrasing is adjusted according to the phonological requirements of the individual languages and thus receives syntax-independent (and in some circumstances non-isomorphic) status.

The main focus in the present chapter is on a group of 'special clitics', namely endoclitics in Degema. There are several issues related to endoclisisis which make such forms a special challenge for the approach developed here. First, it is seemingly a violation of the concept of lexical integrity. Second, it is also problematic if strict modularity is assumed. And third, endoclitics alter the surface order of the string. All three issues will be discussed in more detail throughout the chapter where applicable, with a focus on the concept of the string and its relation to p-structure. It

will be shown that the problematic status of the first two concepts can be resolved via specific assumptions about the string and p-structure, and their position within the grammar architecture.

The chapter is structured as follows: First, a brief introduction into endocclisis is given and its problematic status for the concepts of modularity and lexical integrity is discussed. This is followed by a subsection discussing another possible approach to endocclisis within LFG, *Lexical Sharing*. The introductory section then turns to the string as it is perceived in the LFG architecture in relation to the application of postlexical phonological rules and endocclisis. Section 5.2 gives a brief introduction to the language of Degema before focussing on the factative en(do)clitic in Section 5.3. Section 5.4 analyses the results of the previous sections within the model proposed in this thesis.

5.1.1 Endocclisis

This section introduces one set of ‘special clitics’ belonging to Anderson (2005)’s category 3 (Chapter 4, Section 4.1.1), i.e., clitics which are syntactically anomalous and prosodically deficient at the same time.

Endocclisis is a very rare phenomenon which has been reported for a total of three languages so far: Degema (Kari 2002), Pashto (Tegey 1977), and Udi (Harris 2002). Other reported cases, e.g., European Portuguese, have been reanalysed as mesocclisis (see below). There is also extensive literature on the question if endocclisis indeed exists (see, among others, Anderson 2005, and references therein), a justified doubt given the rarity of the phenomenon, but also motivated by, it seems, an incongruity of endocclisis with main stream ideas of architectures and grammar concepts, among them also the concept of lexical integrity. However, the two phenomena of endocclisis discussed in this thesis (Degema and Pashto (see Chapter 6)) cannot be easily reanalysed. Especially in the case of Pashto, which has been widely discussed in the literature, critical voices often present only part of the data. On the contrary, this thesis assumes that endocclisis exists and presents a solution supported by the modular grammar architecture of LFG.

Endocclisis involves the movement of a clitic *into* the stem of the host. This phenomenon can be caused by different triggers and the respective realizations in the different languages vary. The two cases of endocclisis discussed in this thesis are determined by a specific phonological environment - on the segmental level as well as on the level of prosodic phrasing.¹ This property of endocclisis, the placement within the stem caused by phonological triggers, often leads to a confusion of endocclisis with infixation, which is why the two phenomena are comparatively discussed in the following.

¹Note that Udi endocclitics are seemingly not triggered by phonology (Harris 2002). The relevant clitics of this language might thus belong to Anderson’s type 2 clitics (syntactically anomalous, but not prosodically deficient) and are thus not within the scope of this thesis – although it would be interesting to reanalyse them from a phonological perspective. This is left for further research.

Infixation is part of the *lexical* morphophonological process. It can be triggered by very different factors. McCarthy and Prince (1993), citing French (1988), provide an example from Tagalog, where the affix *um-* is added to a verb stem as a prefix if the stem has an initial vowel. However, if the initial onset is filled by consonantal material, the *um-* is infixated to the right of the onset before the first vowel. The following table exemplifies the paradigm.

<i>stem</i>	<i>um-</i> + <i>stem</i>	<i>translation</i>
aral	um-aral	‘teach’
sulat	s-um-ulat	‘write’
gradwet	gr-um-adwet	‘graduate’

Table 5.1: Paradigm for the Tagalog prefix *um-*.

Other languages include infixation triggered by different processes, e.g., the leftmost stressed foot in Ulwa (McCarthy and Prince 2001) or consonantal patterns in stem templates in Tiene (Hyman 2006). Lahiri and Plank (2008) list these and further relevant triggers for infixation as cases of phonological rearrangement, which they declare to be the “only productive mechanism of getting affixes inside stems (in the right phonological circumstances, with affixes remaining external elsewhere)” (Lahiri and Plank 2008, 40f). This rearrangement is triggered by a *phonological* preference for a certain order for reasons of, e.g., rhythm or because of specific segmental environments.

The question is how infixation can be distinguished from endoclitisis. First of all, infixes have corresponding affixes, while endoclitics have corresponding external clitics.² The infix *adds* information to the stem it modifies, i.e., it changes its semantic or syntactic interpretation. The endoclititic on the other hand is a *separate* syntactic element. This distinction is also reflected by the causes for infixation and endoclitisis respectively. The infix is triggered by *lexical* morphophonological information present in the host. The endoclititic on the other hand is determined by word-external constraints, e.g., the prosodic phrasing or the segmental qualities of the preceding or following syntactic element. Endoclitisis is thus a *postlexical* process, i.e., it applies after the lexical elements have been retrieved from the lexicon, while infixation is a *lexical* process. Nevertheless, the similarities are striking: Both groups are originally affixes/clitics that display infixation/endoclitisis only in a specific phonological environment, and both processes are very rare.

Another closely related phenomenon is *mesoclitisis*, where the clitic is positioned between the stem and an affix. The probably most well-known example for this is the clitic placement in constructions in European Portuguese where the verb is combined with a future tense or a conditional mood. In the following example, the personal pronoun clitic *lo* (‘it’) is inserted between the stem and the future tense inflection (see Anderson (2005), Spencer and Luís (2012), and references therein for

²The criteria for the distinction between affixes and clitics can be found in Chapter 4.

more information on the data).

- (58) faréi → fá -lo -éi ‘I will do it’
 I.will.do do -it -I.will

(van der Leeuw 1995, 32)

Diachronically, this construction derived from (Vulgar) Latin, where the future/conditional was represented by a separate auxiliary following the verb. These two words could be separated by an intervening personal pronoun. In the course of time, this auxiliary was morphologized in the developing Romance languages – however, in the case of European Portuguese, this morphologization process does not seem to be completed as of yet. While French, Italian and Modern Spanish do not allow for a personal pronoun to be positioned between the stem and the suffix, European Portuguese is (still) able to do so (van der Leeuw 1995).

In the case of European Portuguese, mesocclisis can be explained diachronically and the process leading to it is still visible in terms of the strict positioning of the clitic between the stem and the future/conditional suffix. Mesocclisis has also been claimed to apply to Pashto (see Chapter 6). However, while the clitic in European Portuguese is constrained to a morpheme boundary, the same cannot be said for the Pashto data.

Nevertheless, while the diachronic development is of great interest and explains why a certain construction came to be, the synchronic construction has to be learnable by language learners (first and second language acquisition) who are usually unaware of the historical development. As mesocclisis (like endocclisis) is triggered by postlexical processes and leads to similar issues concerning lexical integrity and modularity, the distinction between mesocclisis and endocclisis is not substantial; the only difference is that in the former case, the positioning of the clitic is determined by a formerly diachronic separation of a synchronic morpheme boundary while in the latter case the element moves into the stem itself. This difference is not fundamental in the architecture proposed here as mesocclisis as well as endocclisis is triggered by postlexical factors – the treatment is thus identical.

As a conclusion, endocclisis is defined by its position within the stem (in contrast to mesocclisis, but not infixation) and its determination through postlexical processes (in contrast to infixation, but not mesocclisis). This leads to the following definition:

- (59) **endocclisis**: Caused by postlexical (phonological) triggers, a clitic is moved into the stem of the host.

The concept of endocclisis involves several modules of language and is thus the optimal candidate for interface discussions. It also poses a challenge to linguistic theory as it seemingly violates the concept of lexical integrity which states that each morphologically complete word corresponds to one and only one c-structure node (see Chapter 2, Section 2.3.2).

However, one should carefully distinguish between the different components involved in the generation of endoclitics. The principle of lexical integrity is concerned with the interaction between syntax and morphology and the separation between words and phrases. The endoclitics presented in this thesis, on the other hand, also involve postlexical phonology in p-structure and it is the discussion of this component and the resulting architectural decisions that will allow for the realization that lexical integrity and endoclitisis are not exclusionary concepts, but, on the contrary, are complementary from a specific architectural viewpoint.

Before these issues are discussed in more detail with a concrete case of endoclitisis, the following section will introduce another approach to en(do)clisis proposed in LFG: *Lexical Sharing*.³

5.1.2 Previous approaches to endoclitisis in LFG: Lexical Sharing

An alternative approach to the one proposed in this thesis concerning elements that are contracted with or attached to a host is provided by *Lexical Sharing* (Wescoat 2002, 2005, 2009) which allows for one ‘lexical exponent’ to project to two terminal nodes. The theory of Lexical Sharing and its application to Udi endoclitics will be introduced in this section followed by a discussion of why it is *not* applied to the data presented in this thesis.

In his thesis, Wescoat (2002) discusses (among others) the analysis of English personal pronoun-auxiliary contractions (e.g., *I will – I’ll*). Following Spencer (1991) and Sadler (1997), Wescoat distinguishes between two forms of the reduced auxiliary: The syllabified form ([l]), which can occur with all possible hosts (e.g., personal names) and the unsyllabified form ([l]), which is restricted to the combination of the auxiliary with a personal pronoun. This restriction leads to the assumption that the unsyllabified reduced auxiliary is in fact part of the inflectional inventory of the personal pronoun. However, this analysis is problematic in that the personal pronoun and the auxiliary instantiate two separate syntactic terminal nodes, but correspond to only one prosodic (lexical) word.

This stands in contrast to the general assumption where words/formatives instantiate to atomic phrases/terminal nodes in a one-to-one correlation; a relationship which is also encoded in the principle of lexical integrity. Di Sciullo and Williams (1987, 107) follow this assumption and propose a solution where *I* and *will* are syntactically analysed before they are contracted in a postlexical process. Wescoat objects to this analysis on the basis of the argumentation that *I’ll* must be a lexical exponent on the one hand (based on the patterns of syllabification mentioned above), but instantiates two syntactic nodes on the other hand. This process is demonstrated in Figure 5.1 (cf. Wescoat 2005, ex.20).

³A further approach to endoclitisis within LFG, Lowe (forthcoming), will be discussed in the next chapter, as it specifically concerns Pashto second position clitics.

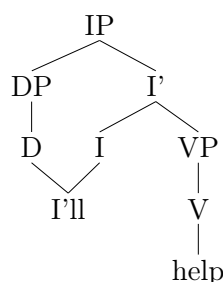


Figure 5.1: Two syntactic nodes corresponding to one lexical exponent.

However, Figure 5.1 does not represent a well-formed c-structure tree as it was defined by Partee et al. (1990) (see Section 5.1.3.1). Wescoat thus proposes to distinguish syntactic nodes that indicate phrases from syntactic nodes that contain word forms, because otherwise “the constraints on the domination relation that result in mother uniqueness affect nodes representing phrases and nodes representing words in exactly the same way” (Wescoat 2002, 12). In order to separate syntactic nodes from word forms, he defines a new structure, the *lexical-sharing tree*, separating phasal structure from lexical parts. In the following, this proposal is explained in more detail (cf. Wescoat 2005).

1. Sever the words off a traditional c-structure tree and transfer them into a separate *l(lexical)-structure* in linear order — later match the words in l-structure against the lexicon. Consequently, the former pre-terminal nodes (e.g., N, A, D), which Wescoat calls ‘atomic constituents’, are now the terminal nodes of the new tree.
2. The relation between the atomic constituents and the words in l-structure is established by a correspondence mapping λ . A word w in l-structure can be viewed as the ‘atomic constituent X’s lexical exponent’; it *instantiates* X ($\lambda(X)=w$). λ is a many-to-one correspondence, in that the number of syntactic atoms can exceed the number of lexical exponents.

In addition to the traditional phrase structure rules ($S \rightarrow NP$, etc.), lexical sharing requires a second type of rule specifying the atomic constituents which a word may instantiate. For *I’ll* in the above example, such a lexical exponence rule is shown in (60), where the lexical form is given on the left-hand side and a string of the related atomic constituents on the right-hand side.

$$(60) \text{ I'll} \leftarrow N V$$

Lexical instantiation rules like the one in (60) can thus be used to restrict the contracted non-syllabic auxiliary to the combination with personal pronouns.

The main idea behind this proposal is that λ separates the formatives from the atomic constituents and thus from the domination constraints of a phrase structure

tree. As a result, *two* atomic constituents can be mapped to *one* word. Figure 5.2 demonstrates this relationship, where the λ correspondence between phrase structure and l-structure is indicated by arrows.

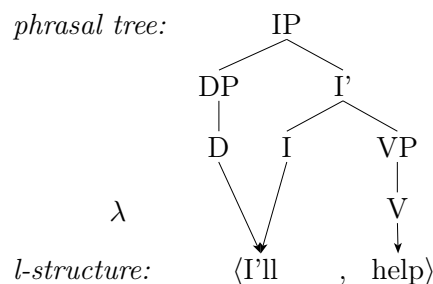


Figure 5.2: Lexical Sharing tree (Wescoat 2005, ex.22).

This proposal is restricted to a number of constraints, e.g., the atomic constituents and the corresponding lexical exponents in l-structure must be strictly parallel in the linear order. Furthermore, Lexical Sharing can only apply to two (or more) adjacent constituents: Lexical Sharing across several non-involved constituents is not permitted.

Concerning the principle of lexical integrity, Lexical Sharing is conform with the assumption that no syntactic rule can refer to elements of morphological structure; the λ correspondence maps constituents to unanalyzed words and is not aware of the word's internal structure. However, the notion that each word corresponds to one c-structure node has to be adjusted. Wescoat (2002, 22) coins the term 'homomorphic' lexical integrity (in contrast to 'isomorphic' lexical integrity) which states that the precedence relation between two terminal nodes must be identical to the precedence relation of the corresponding lexical exponents in l-structure, and that terminal nodes sharing a lexical item must be adjacent.

5.1.2.1 Lexical Sharing and Udi endocclisis

Lexical Sharing has been applied to a number of clitic(like) phenomena (Broadwell 2008, Alsina 2010) and Wescoat (2002, 46) himself states that "cliticization could be regarded as a lexical rule that takes the host as input, attaches the clitic as an affix, and causes the result to instantiate an additional atomic phrase" – on the other hand, he also notes (Wescoat 2009, footnote 6) that Lexical Sharing might not be applicable to all types of clitics, but probably at least to those described as *simple* clitics by Zwicky (1977).

One of the phenomena discussed by Wescoat is Udi endocclisis where he applies Lexical Sharing to Udi endocclitics followed by an OT analysis to explain the occurrence of the clitic within the stem (Wescoat 2009). He views the Udi person

markers as “instantiation-altering morphemes”⁴, that is, a word containing a person marker instantiates two terminal nodes as encoded in the corresponding lexical exponence rule. Example (61) shows such a lexical exponence rule, where the Udi verb *bey-al* ‘will watch’ is combined with an enclitic *le* (3Sg) (PM stands for person marker and \downarrow is an abbreviation for $\phi(\lambda(*))$; $\uparrow = \downarrow$ has been added to clarify the functional annotation):

$$(61) \quad \text{bey-al-}\underline{\text{le}} \leftarrow \begin{array}{l} \text{V} \\ \uparrow = \downarrow \\ (\downarrow \text{ PRED}) = \text{'WATCH} < (\downarrow \text{ SUBJ}), (\downarrow \text{ OBJ}) > \text{' } \\ (\downarrow \text{ TNS}) = \text{FUT} \\ \downarrow = \downarrow \end{array} \quad \begin{array}{l} \text{PM} \\ \uparrow = \downarrow \\ (\uparrow \text{ SUBJ}) = \downarrow \\ (\downarrow \text{ PERS}) = 3 \\ (\downarrow \text{ NMB}) = \text{SG} \end{array}$$

(Wescoat 2009, ex.49)

These sequences of lexical exponence rules thus attach clitics like suffixes to all possible hosts in the lexicon. The corresponding tree fragment is given in Figure 5.3 (cf. Wescoat 2009, ex.50).

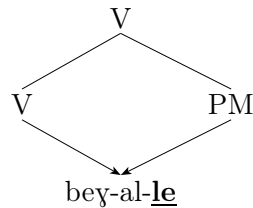


Figure 5.3: Lexical Sharing applied to *bey-al-le* (watch-FUTII=3SG).

The unit resulting from the lexical exponence rules in (61) instantiates two (and possibly more) terminal nodes. However, while the lexical exponence rules above account for the person marker as an enclitic, they do not per se explain the occurrence of endocclisis. In order to account for the different positions that the PM can occupy, Wescoat has to rely on additional formal power. Based on the optimality constraints proposed in Harris (2002), Wescoat thus combines Optimality Theory with the concept of Lexical Sharing to explain the cases of endocclisis (see Wescoat (2009) for details).

Such additional power would be needed as well if Lexical Sharing was applied to special (endo)clitics that are clearly triggered by phonological constraints. However, as has been stated in Chapter 4, Section 4.3.2, an OT approach is not necessarily a desirable solution; especially not if there is a solution possible if the right grammar architecture is assumed. And there are other issues with the Lexical Sharing approach to (endo)clitics: First, each word to which a clitic can attach will have

⁴These clitics are part of an *instantiation-altering morphology* which Wescoat assumes in addition to inflectional and derivational morphology.

to be listed *with* the clitic. This will lead to a serious proliferation of the lexicon, especially if the clitics are promiscuous, i.e., if there is a large number of potential hosts, or if there are clitic clusters (as it is the case with Pashto, see Chapter 6). In the approach presented here, this potentially infinite enlargement of the lexicon is replaced by a (finite) set of abstract rules accounting for endocclisis in p-structure.

Second, Lexical Sharing relies on a modified version of the principle of lexical integrity, the *homomorphic* lexical integrity which states that the precedence relation between two terminal nodes must be identical to the precedence relation of the corresponding lexical exponents, thus allowing for a one-to-many relation between the lexical exponent and the corresponding terminal node(s).⁵ In contrast, lexical integrity is still intact in the approach presented here.

Nevertheless it can be argued that Udi endoclitics are only partly comparable with the clitic phenomena presented in this thesis, as it is not clear in how far phonological factors play a role with Udi endocclisis.⁶ For phenomena that are clearly triggered by phonological/prosodic circumstances, however, the application of Lexical Sharing is rejected in favour of the approach presented here, which respects the principle of lexical integrity and the concept of modularity, preserves the integrity of c-structure and at the same time explains a complex phenomena without the application of additional formal power outside of the grammatical modules assumed in linguistic analysis.

However, before the approach presented here can be explained in more detail, one more relevant concept has to be introduced: The theory of the string and its role within the architecture.

5.1.3 String and p-structure in the LFG architecture

The previous chapter introduced postlexical phonology as a set of rewrite rules and constraints that apply to a given input within p-structure. This included a specific view of the grammar, where p-structure and c-structure were positioned on the respective sides of the string. Figure 5.4 represents this underlying assumption of the *parallel* structures.

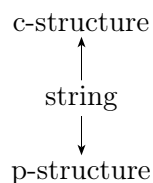


Figure 5.4: C-structure and p-structure position with respect to the string.

⁵Strictly speaking, the precedence relation between the host and the clitic is no longer given in the case of endocclisis.

⁶This is, however, not true for Wescoat's initial motivation for the Lexical Sharing approach, the English personal pronoun-auxiliary contraction (Wescoat 2002, see also Section 5.1.2) – this topic is left for further research.

In the case of Swabian pronoun clitics the output of p-structure was parallel to the string analysed by the syntax. However, the question arises how to deal with cases where the linear order of the p-structure output is altered *only* because of p-structure internal processes, i.e., postlexical phonology – and where the alternation is unmotivated by syntactic requirements. Such a change of linear order occurs with the cases of endoclisisis presented in this thesis, and it requires a more detailed analysis of the concept of the string and how it is positioned with respect to p-structure (and the postlexical phonological rules).

In general, the syntactically unparsed, but tokenized string is simply taken to be mapped to c-structure via the relation π (Kaplan 1987, Asudeh 2006), where c-structure represents the ‘linear order and hierarchical structure’ of that string (Dalrymple 2001, Asudeh et al. 2013). This concept of the string is described in more detail in Asudeh (2009) who defines the (s-)string as a ‘representation of linear phonology’, thus assuming phonological and syntactic linear order to be parallelly represented within the string. A similar proposal is made by Dalrymple and Mycock (2011) (Chapter 2, Section 2.6) who assume that the string is at the heart of the interface, incorporating (at least) two representations: the p-string which represents the phonological form of the string, and the s-string, which represents the syntactic side of the string. Both representations are interfaced by a multi-dimensional lexicon, which includes information on the phonological and syntactic properties of a lexical entry. This two-fold perspective of the string implies a parallelity of the p- and s-string, an implication which is made explicit by Lowe (forthcoming) (see Chapter 6, Section 6.3.2). Another recent analysis has adjusted the notion of linear order in that second position clitics were allowed to be moved in order to correctly analyse the associated syntactic information (Bögel et al. (2010), following Halpern (1995)), which implies a non-parallelity between the two sides of the string.

This discussion about p-string and s-string also implies the question on where postlexical phonological processes are posited. If these are responsible for the rearrangement of elements in a string, then there are two options⁷:

1. The postlexical phonological processes are (in close relation with the lexicon) an external *interpreter* device. In this case, the elements given in a string are interpreted by looking up possible disambiguations through phonological rules before identifying an element via lexical look-up. This approach keeps the p-string/s-string parallelity intact.
2. The postlexical phonological processes have a *translatory* function between the two versions of the string, in that the p-string, which contains the surface variation, is ‘translated’ into the s-string by first disentangling the variation

⁷A third part-option proposed by Lowe (forthcoming) is rather placed within the syntactic component and assumes a non-parallelity between s-string and the linear order of c-structure. See Chapter 6, Section 6.3.2 for some discussion.

Figure 5.5 has two possible c-structure representations. An amended version of option a) has been assumed under the proposal of *Lexical Sharing* (see Section 5.1.2), while option b) is the underlying assumption made in Bögel (2010)¹⁰. Both c-structure options preserve the parallelity between p- and s-string as proposed by Dalrymple and Mycock (2011). The s-string under this view is thus the linear representation of the incoming speech signal. The postlexical component is an external resource which is used together with the lexicon to syntactically identify the elements in the string.

However, there are several issues with this approach. First of all, according to Partee et al. (1990, 439), neither Figure 5.5 (a) nor (b) represent a well-formed (c-structure) tree. Partee et al. define a linguistic tree as representing three types of information about syntactic structure: a) the hierarchical grouping of the constituents, b) the grammatical type of each constituent, and c) the linear order of the constituents. Furthermore, the tree is constrained in that branches are not allowed to cross each other and every node has at most one node dominating it. These two constraints on dominance and precedence are expressed by the *Nontangling Condition*, given in (63).

- (63) **Non-Tangling Condition:** In any well-formed constituent structure tree, for any nodes x and y , if x precedes y , then all nodes dominated by x precede all nodes dominated by y . (Partee et al. 1990, 442)

The restriction on linear order is violated in Figure 5.5 (a) by the fact that the terminal node dominated by V should precede the terminal node dominated by CL – which it doesn't. The tree in Figure 5.5 (b), on the other hand, seemingly does not violate this principle of linear order and dominance, but does not clearly represent the grammatical type. It also implies an integration of the clitic somewhere in the lexicon and thus simply shifts the problem from one module to another. The integration of clitics within the lexicon requires a) considerable more lexical space, especially if the clitics attach to several categories and b) a motivation of the clitic's treatment within the lexical component, even though its triggers are found outside of the lexical form. The representation given in 5.5 (b) is thus rather a deceptive packaging than an alternative solution.

The restriction on dominance is closely related to the principle of lexical integrity, which demands that each terminal node dominates only one morphologically complete word. This is problematic for the approach presented in this section, as the representation of verb and endoclititic as one terminal node within the tree clearly violates the notion.

Another issue is the embedding of the clitic information *within* the verb's f-structure information. This may not be a problem with clitics attaching to the verb as the functional information of the verb usually reaches clausal scope and thus also

¹⁰Bögel (2010) is an (incomplete) preversion to Chapter 6 and is thus not further discussed in this thesis.

allows for the clitic's functional information to reach the top-level in f-structure. However, some clitics may be embedded in lower constituents, e.g., a subject NP, but their information may be important for the clausal functional level.¹¹ This non-isomorphic relationship of the syntactic (functional) position and the prosodically rearranged 'overt' position is a tension that cannot be easily resolved with the proposal presented in this section.

In addition to these issues, there is also a problem from the perspective of modularity. The analysis under one terminal node can only be explained from the perspective of *comprehension*, because only then is the postlexical phonological information that caused the endoclysis available before lexical and syntactic analysis; i.e., the trigger for endoclysis is available before the clitic and its host are analysed by syntax. However, it is difficult to explain how the endoclititic came to be within the verb if the opposite direction (*production*) is considered. In this case, the trigger for the postlexical phonological process of endoclysis is 'not available' as of yet, but will only be available at a 'later' stage after the syntactic analysis. In order to keep the p-string and the s-string parallel, however, the endoclititic *has to be present* in the s-string. Endoclysis within c-structure is then an anticipation of possible-phonology-in-the-future and consequently a violation of modularity.

5.1.3.2 Postlexical phonological rules in a translation related function

Another approach to the issue is the assumption that syntactic linear order and postlexical phonological variation may differ. In that case, the p-string and the s-string are not parallel and the postlexical phonological rules take up a translation-related function between the two strings. In the c-structure, the clitic is placed in its 'original' position while the endoclititic variation triggered by postlexical phonological constraints applies in the p-structure, separated from c-structure.

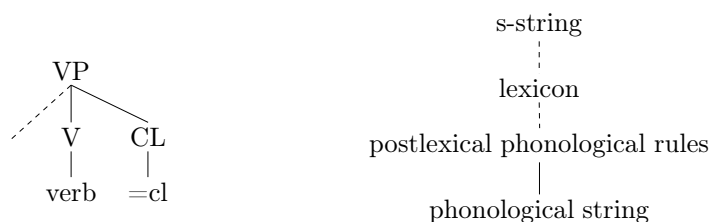


Figure 5.6: Partial syntactic tree and architecture for s-string \neq p-string.

In Figure 5.6, the s-string is not parallel to the p-string. In this approach, the clitic occupies its own terminal node, following its host in an *enclitic* position, although the clitic can also be placed further away from the host, e.g., as it is often the case with

¹¹As it may be the case with, e.g., second position clitics (see Bögel et al. (2010) and also footnote 12 below).

second position clitics. The postlexical phonological rules in p-structure, which are responsible for the positioning of the clitic within the host, are situated between the s- and the p-string. They have a translation related function in that they decompose the incoming speech signal before the lexical analysis according to the language's postlexical phonological rules. As a result, p- and s-string are parallel most of the time, but might differ in the case of postlexical phonological intervention. From this perspective, the principle of lexical integrity is not violated, as the clitic occupies its own terminal node. Modularity is also maintained, because the clitic's placement according to phonological constraints is restricted to p-structure while its syntactic placement is determined by syntacto-functional considerations. Furthermore, a syntactic analysis in general is much easier, as the clitic is not locked into another syntactic (and possibly functionally completely unrelated) element.

However, there are also drawbacks to this view, e.g., if the 'original' position of the clitic is unknown, as it is the case with some second position clitics: If they do not have a corresponding full form, it is difficult to motivate their 'movement' through the application of postlexical phonological rules. Nevertheless, for most second position clitics, a syntactic position can be inferred, e.g., by analysing functional scope.¹²

As the approach assuming parallelism of the strings discussed in Section 5.1.3.1 involves problematic consequences for lexical integrity, modularity, syntactic dominance and functional interpretation, and furthermore, is less flexible with respect to postlexical phonological variation, this thesis proposes that postlexical phonological processes occupy a translation related function between p- and s-string. As a consequence, s- and p-string are isomorphic most of the time, but, under certain prosodic/phonological conditions, they do not have to be. These conclusions are neatly expressed in a quote by Lahiri and Plank (2008, 45):

It's (prosodic) phonology [...] that masterminds overt variation.

The following section formalizes this approach via the analysis of endocclisis in the language of Degema, and shows how modularity and lexical integrity can be maintained on the basis of two assumptions:

1. The linear order of the "speech signal" (the p-string) is not necessarily equal to the linear order of the s(yntactic)-string
2. Assuming modularity, the postlexical phonological rules that are responsible for endocclisis refer to *prosodic* (and not syntactic) boundaries to determine their domain of application.

¹²The Serbian/Croatian/Bosnian clitics discussed in Bögel et al. (2010), for example, are syntactically analysed in the first position of the sentence, but prosodically expressed in the second position, following the concepts of *prosodic inversion* proposed by Halpern (1995). The choice of the first syntactic position is motivated by a) the then-available sentential functional scope, b) by an unwillingness of each clitic to appear anywhere else in the sentence, and c) by their restriction to the first position after the complementizer in subordinate clauses.

5.2 Degema: Some phonological background

Degema is a Delta-Edoid language, spoken in the Rivers State region of Southern Nigeria. So far, no standard version of Degema has emerged, but there are two dialects: Usokun and Atala, spoken by an estimated 11,000 speakers each. The main focus of this chapter lies on the Usokun dialect. The following examples, their orthographic representation and the gloss are mostly taken from the works of E. Kari (2002, 2003, 2004, 2007, 2012).

From the phonological viewpoint, Degema has several phonological aspects which are crucial for the discussion of endoclysis in Section 5.3.2. Thus, this section briefly introduces Degema syllable structure, vowel harmony and lexical tone.

1. Syllable structure

Degema has four basic syllable types: V, VC, CV and CVC (in short: (C)V(C)). While all consonants can occur in word-initial and word-medial position, only few can occur word-finally. Consonant clusters only occur word-initially and word-medially, and are the result of deleting an intervening vowel, thus CCV is derived from CVCV. In connected speech, consonants will resyllabify to a following word starting with a vowel. Thus, the prevalent syllable structure is CV or CVC (cf. Kari 2004, 378ff.).

2. Vowel harmony

Degema has ten vowels: /i, ɪ, e, ε, a, ə, ɔ, o, ʊ, u/, which can be divided into two sets: One set with an advanced tongue root (phonetic feature +ATR) /i, u, e, o, ə/ and one set with a retracted tongue root (-ATR) /ɪ, ʊ, ε, ɔ, a/ (Fulop et al. 1998). Vowels in simple words are exclusively drawn from one of the sets and are thus subject to vowel harmony. The type of vowel given in the stem will also determine the vowel of any affix attached to the stem. Vowel harmony also spreads to clitics, possessives, some object pronouns and the negative adverb (Kari 2007).

3. Lexical tone

Degema is a tone language, which can distinguish between segmentally and categorically identical lexical entries via lexical tone and thus requires an elaborate analysis of all linguistic aspects of a lexical entry. Degema has two basic tones, H(igh) (x̂), L(ow) (x̃, or unmarked) and a downstepped high tone (indicated by ↓x), which applies if two adjacent vowels carry a high tone each – a weakened variant of the *Obligatory Contour Principle* (OCP) which states that adjacent like tones are banned (Leben (1973), see also Gussenhoven (2004) and references therein). Tones are marked on the syllables of a string.

As reported by Kari (2004), there is no evidence of contour tones, i.e., two tones do not share one syllabic host. Degema can thus be described as a *register tone language* (Pike 1967b), as each syllable has one level tone. There are also no long vowels in Degema (the word for ‘yes’ being an exception). Sequences of identical

vowels occur only in a handful of words and have a high - downstepped high ($\acute{V}\downarrow V$) tone pattern.

The distinctive pitch level carried by each syllable of a word is crucial to the correct interpretation. In Figure 6.2, the segmentally identical word /ugo/ is shown in two of its three possible combinations of tone (data from UCLA Phonetics Lab Archive (2007)).

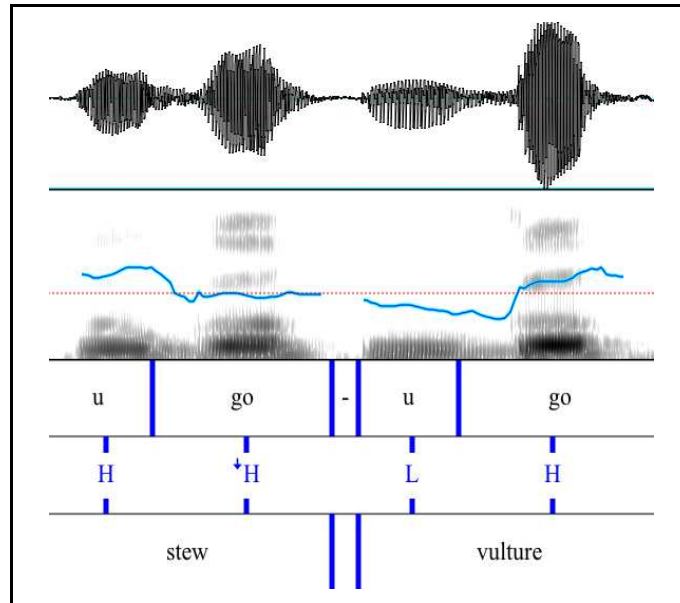


Figure 5.7: Speech signal for $\acute{u}g\grave{o}$ ‘a kind of stew’ and $\grave{u}g\acute{o}$ ‘vulture’ .

Apart from the two possibilities shown in Figure 6.2, there is (for the Usokun dialect) a third sequence of tones for /ugo/: $\acute{u}g\grave{o}$, high-low, which is connected to the concept of ‘butterfly’.

The approach to use identical segments in combination with a distinct tone sequence to distinguish differences in meaning shows a different concept of pitch assignment in comparison with the languages discussed in the previous chapters: While German and English exhibit *postlexical* pitch assignment and lexical stress, Degema features *lexical* pitch assignment, but no lexical stress (see Ladd (2008, 164ff.) for an overview on these distinctions). This requires some adjustment to the phonological form of the lexicon as it was introduced in Chapter 3. Thus, a lexical entry in Degema must not only contain information on segments and syllable structure, but requires a description of lexical tone as well. Table 5.2 shows an extension of the lexicon, distinguishing between concepts and the related s(yntactic)-forms and p(honological)-forms, where the respective p-form can contribute to a quick disambiguation of the segmentally identical strings by means of lexical tone.

concept	s-form	p-form
PALM-NUT VULTURE	ugo N (\uparrow PRED) = ‘ugo’	< ùgó > SEGMENTS /u g o/ METRICAL FRAME $(\sigma\sigma)_\omega$ LEXICAL TONE LH
BUTTERFLY	ugo N (\uparrow PRED) = ‘ugo’	< úgò > LEXICAL TONE HL
A KIND OF STEW	ugo N (\uparrow PRED) = ‘ugo’	< úg [↓] o > LEXICAL TONE H [↓] H

Table 5.2: Lexical entries for ùgó ‘vulture’, úgò ‘butterfly’ and úg[↓]o ‘a kind of stew’. The p-form entry for Degema consists of three aspects. First, the segments of the string are listed; in Figure 5.2, the segments for all three entries are identical. Second, the metrical frame of the entry states the number of syllables among which the segments are distributed. Furthermore, the brackets around the two syllables show that the lexical entry is a prosodically complete ‘word’, meaning that the lexical entry does not depend on a host (as, e.g., the factative clitic does, see below in Section 5.3). Third, the lexical entries of Degema feature lexical tone, which is also distributed among the syllables. In the case of the segmentally identical word *ugo*, the only difference between the three lexical entries is lexical tone. The correct concept/s-form belonging to a speech signal section is thus only available, if the lexicon allows for an identification of phonological features (especially lexical tone) as well.

From the comprehension perspective, the information on tone is provided by the speech signal itself and stored in the p-diagram (Bögel 2012, Bögel 2013). Figure 5.8 shows how the information on tone is displayed in the p-diagram.

LEX. TONE	H	↓H	...	LEX. TONE	L	H	...
SEGMENTS	[u]	[go]	...	SEGMENTS	[u]	[go]	...
VECTORINDEX	S ₁	S ₂	...	VECTORINDEX	S ₁	S ₂	...

Figure 5.8: p-diagram representations of *ugo*: ‘stew’ (left) vs. ‘vulture’ (right).

The information stored under LEXICAL TONE in Figure 5.8 can subsequently be used for the matching between the speech signal and the respective syntactic item, thus providing the correct syntactic information for further processing. From the perspective of production (from meaning to form), this information has to be transferred from the lexicon during the *transfer of vocabulary* as it was introduced in Chapter 4.

While the above example shows how multi-dimensional lexical entries can aid in the disambiguation process on the word level, this concept of the lexicon is also vital for an analysis of notoriously difficult phenomena like endoclititics.

5.3 The factative clitic

Degema has several clitics; of these, the factative clitic is of special interest, as it can appear as an enclitic in some contexts, but as an endoclititic in other contexts. The factative clitic attaches to verbs and object pronouns¹³ and thus fulfills CRITERION A (Zwicky and Pullum (1983), see Chapter 4, Section 4.1.1) which states that clitics exhibit a low degree of selection to their hosts in comparison with affixes. Other criteria, e.g., the lack of arbitrary gaps or specific idiosyncrasies also point towards the categorisation of the factative as a clitic and not an affix. The factative aspect

is used to denote a fact, which may be a dynamic situation that has already been completed or a state that once existed or still exists at the present time.

(Rose (2014), following Jenewari (1980, 133))¹⁴

The factative clitic consists of an underspecified vowel which copies all features of the vowel (i.e., the complete vowel) in the host's last syllable as part of a process of vowel harmony, followed by *n*: *Vn*. The clitic's realisation depends on

- a. the phonological environment (consonant vs. vowel) of the host's last segment and (if present) the following word's first segment.
- b. its medial vs. final position in an intonational phrase (*ι*).¹⁵

Note that due to its factative meaning the clitic is *not* realized in a question, within a negative clause, or in the presence of a clitic indicating the future. However, these aspects are not determined by phonological factors and are thus left out of the current discussion as their analysis would go beyond the scope of this dissertation. In the following sections, the different realisations of the clitic will be demonstrated with a number of examples from Kari (2004).

¹³It could be claimed that the object pronoun is actually a clitic as well, as the pronoun's vowel quality can be determined by the vowel of the preceding verb stem; i.e., object pronouns in the first person singular and the second and third person plural are subject to vowel harmony. The process of vowel harmony is usually assumed to apply within the prosodic word domain in which case the object pronoun should be interpreted as a clitic attaching to the verb. On the other hand, all of these pronouns can be used in isolation, e.g., as an answer to a question (Kari 2004, 251) – a characteristic, clitics usually do not display, as they are depending on a host. But even if the object pronoun is interpreted as a clitic, this will not transform the factative aspect into an affix, as it attaches *after* the object pronoun, and thus fulfills Zwicky and Pullum's CRITERION F which states that clitics can attach to material already containing clitics, but affixes cannot. Given these reasons, it can thus be stated that the factative aspect is indeed a clitic, no matter if the pronoun is viewed as independent prosodic word or as a clitic itself.

¹⁴Kari prefers the notion *factative*, because it “marks past in dynamic verbs but past/non-past in stative verbs. Given this situation, one can really not describe factative as perfective, since in stative verbs factative could have a non-past or timeless meaning/interpretation.” (Rose 2014, footnote 4).

¹⁵Note that the initial position never applies, as the factative is an *enclitic*, which is also syntactically *always* placed after its host.

5.3.1 Enclisis in the medial position of the intonational phrase

In the medial ι -position, the underspecified vowel is never realised. The realisation of the n is determined by its phonological environment in that it depends on the nature of the last segment of the host and the first segment of the following word. Consequently, there are four possible combinations: 1. $V=FC$ C, 2. $V=FC$ V, 3. $C=FC$ V, and 4. $C=FC$ C.¹⁶ In the following, these possible combinations and the resulting realizations of the clitic are discussed.

1. Phonological environment: **Vowel–Factative–Consonant** ($xxV=n$ Cxx)

- (64) Breno o=síré tǎ=**n** mú éki
 Breno 3Sg=run go=FC to market
 ‘Breno ran to the market.’ (Kari 2004, 115)

2. Phonological environment: **Vowel–Factative–Vowel** ($xxV=n$ Vxx)

- (65) Ení ɓól-ám ójzí yɔ i=díyómósé=**n** ávom ɓáaw
 we hold-GER thief DEF 3Sg=sweeten=FC inside their
 ‘It pleased them that we caught the thief.’ (Kari 2004, 50)

In examples (64) and (65), the host ends in a vowel. In these cases, the clitic is always realised as n . This is also true if the clitic follows a host ending in a consonant and precedes a vowel-initial element, as it is the case in example (66).

3. Phonological environment: **Consonant–Factative–Vowel** ($xxC=n$ Vxx)

- (66) Uɓuwan i=kél=**n** úsóm yɔ
 salt 3Sg=be_more_than=FC soup DEF
 ‘Salt is more than the soup.’ (Kari 2004, 153)

There is, however, one exception to this pattern: In cases, where the final consonant is n , one of two adjacent identical consonants is elided.

- (67) Mǐ=nyán ák[↓]i
 1Sg=have.FC cooking pot
 ‘I have a cooking pot.’ (Kari 2004, 159)

The clitic is also suppressed between two consonants ((68)).

4. Phonological environment: **Consonant–Factative–Consonant** ($xxC \emptyset$ Cxx)

¹⁶V stands for ‘vowel’ and C for ‘consonant’. Following Kari (2012), FC is used to indicate the factative clitic within the examples. All endoclititic examples are annotated with subscripts as introduced in Section 5.1.3.1.

- (68) E=yáw mǫ́ ínwíny útany
 3Pl=take.FC from body tree
 ‘They got it from a tree.’ (Kari 2004, 200)

In this specific environment, the factative enclitic is not realised which is consistent with the general rejection of complex consonant clusters mentioned in Section 5.2. Thus, the generalisation can be established that the factative clitic is realised as *n* in medial intonational phrase position, except when preceded and followed by a consonant or if the preceding consonant is *n*.

5.3.2 En(do)clisis at the final position of the intonational phrase

The pattern described in the previous section changes if the clitic appears at the final position of an intonational phrase. Example (69) shows the clitic following a host ending in a vowel.

1. Phonological environment: **Vowel–Factative (xxV=Vn)**

- (69) O=síré=↓en
 3Sg=run=FC
 ‘(S)he ran.’ (Kari 2004, 72)

If the factative clitic follows a host ending in a vowel, the underspecified vowel of the clitic copies the features of the host’s last vowel. In (69), *Vn* is thus realized as *en*. The presence of the factative vowel is indicated via a downstepped high tone, as two adjacent high tones would violate the *Obligatory Contour Principle* (cf. Section 5.2). The two tones each belong to a separate syllable nucleus.

Endoclisis, finally, occurs if the clitic is in the final position following a host that ends in a consonant.

2. Phonological environment: **Consonant–Factative (xxVVC) → endoclisis**

- (70) O=ǫ́ó=↓o=l
 3Sg=hold₁=FC=hold₂
 ‘(S)he held (a cloth).’ (Kari 2004, 72)

As in (69), the underspecified vowel first copies all the features of the host’s last vowel; it is then moved into the last syllable via *metathesis*, i.e., by switching position with the last consonant of the host, forming a *separate* syllable (Kari 2002).¹⁷

Metathesis is generally defined as the “reordering of phonemes” which “often accompanies affixation” (Spencer 2007, 138). Strictly speaking this only involves

¹⁷A possible alternative explanation to endoclisis would be that the factative clitic in final *ι*-position causes the vowel of its host to be lengthened and to adapt two different tones. However, this is not conform with the general phonological findings of Degema, which neither allow for vowel lengthening nor contour tones.

CV/CC alternations which are quite common across languages.¹⁸ However, as Kari (2003, 100) notes, there are also cases where the elements involved in metathesis are ‘bigger’ than a single phoneme, e.g., morphemes and clitics consisting of more than one phoneme. Metathesis of syllables with other elements in the context of infixation is also mentioned in Lahiri and Plank (2008, 42ff) who give an example of Tagalog perfective verbs, where the prefix *um-* is infixated after an initial consonant.¹⁹ This is similar to the metathesis found in Degema, although in this case, it involves a clitic and takes place at the right edge of the verb, moving the factative clitic to the position preceding the final consonant. This metathesis with the final consonant of the host can be found in other languages as well, albeit mostly on the level of infixation (e.g., Latin) (cf. Lahiri and Plank 2008, 42). However, interestingly, Udi endoclisism is also formed by moving the respective element into the position before the final consonant of the host.

Following the process of metathesis, the *n* is deleted in order to avoid illegal consonant clustering which is consistent with the deletion in the context $xxC \emptyset Cxx$ ((68)) and the general impossibility of word-final consonant clusters in Degema. After vowel harmony has applied, the second high tone is downstepped as required by the OCP. The result is a changed tone pattern H^+H as found in examples (69) and (70), which contrasts with the factative clitic in medial *ι*-position where the tone pattern is maintained. Table 5.3 gives a short overview on the processes involved in Degema endoclisism.

1.	underlying form	ból+Vn
		↓
2.	metathesis	bóVnl
		↓
3.	n-deletion	bóVl
		↓
4.	vowel harmony	bóól
		↓
5.	high tone downstep	bó [↓] ol

Table 5.3: Endoclisism in Degema

Note that step 2–4 are variable to a certain extent, i.e., they can be switched around without any impact on the surface form. Further research in the language would be required to see if these phonological processes apply in a specific order or if they are indeed parallel; as this is not essential to the current discussion, it is left for future research.

The examples (69) and (70) show the respective form of the factative clitic in the final position of a sentence. However, as the following examples show, the realization

¹⁸See <http://metathesisinlanguage.osu.edu/> for an overview and a language database

¹⁹See also Crowhurst (1998) and references therein, and Section 5.1.1, Table 5.1.

of the underspecified vowel carrying a downstepped high tone and the process of endoclisisis is also available at the end of subordinate clauses.

- (71) [E=kótú mé=[↓]en] dɔ mí=meme
 3Pl=call me=FC but 1Sg.NEG=answer
 ‘They called me but I didn’t answer.’ (Kari 2004, 139)

- (72) Owéy [nú ábɔ ɪ=vúwó=[↓]o=y] yɔ o=yí=t[↓]e
 person that hands 3Pl=be_dirty₁=FC=be_dirty₂ DEF 3Sg=come=PE
 ‘The person whose hands are dirty has come.’ (Kari 2004, 202)

The following figure shows an oscillogram and a spectrogram of a sentence containing endo- and enclisis annotated with Praat (Boersma and Weenink 2013).

- (73) [ómóʃítám jɔ ɔ=sɔ=[↓]ɔ=l] [ɔ=kírí béné=ⁿ úkpá]
 girl DEF 3Sg=jump₁=FC=jump₂ 3Sg=also play=FC dance
 ‘The girl jumped and danced.’ (Kari, p.c.)

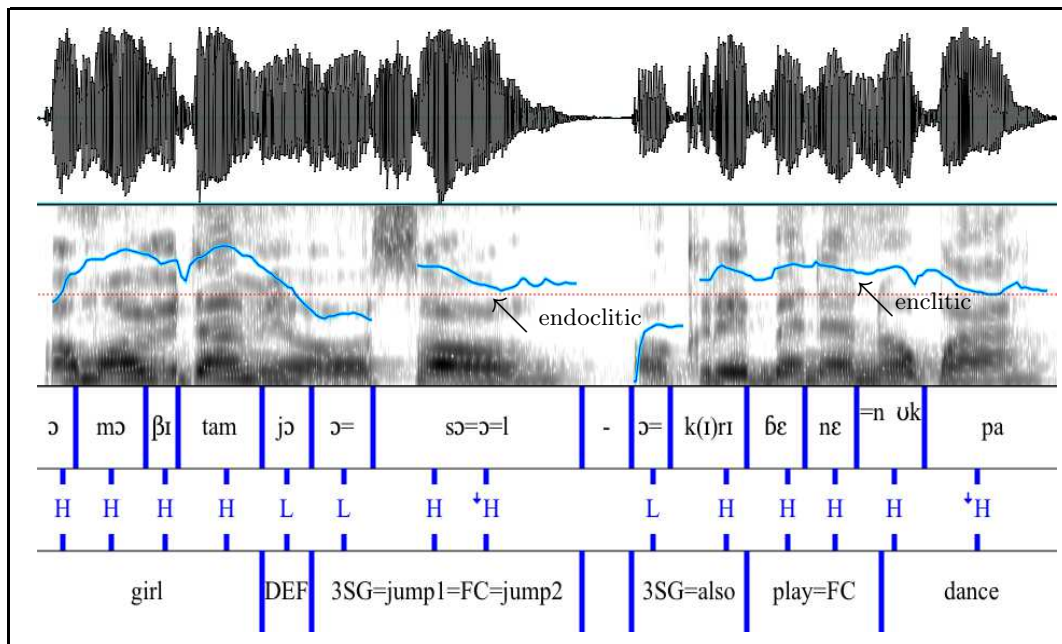


Figure 5.9: Speech signal for (73): ‘The girl jumped and danced.’

Figure 5.9 consists of two intonational phrases. The first intonational phrase is terminated by the factative clitic. Since the host (*sól*) ends in a consonant, the clitic moves into the host and the *n* is deleted. The clitic’s vowel is marked by a downstepped High tone, as indicated by the first arrow in the speech signal, where

the pitch clearly moves from a high pitch mark to a lower one. The second part of the sentence contains a factative clitic in a medial intonational phrase position. Here, the vowel is not realised and no downstep occurs; the factative clitic is represented by *n*, as indicated by the second arrow in the speech signal.

As in the examples before, the intonational phrases in (73) are isomorphic with the syntactic clauses which leads to an interesting question: Is it really the position at the right edge of a prosodic domain that is responsible for the triggering of endoclisisis, or is it rather the end of a syntactic clause?

5.3.3 Prosodic phrase or syntactic clause

The previous examples of the factative in the final phrase position don't make a clear statement on the question if the trigger for endoclisisis is indeed the intonational phrase (as it was presumed up to this point) or rather the syntactic clause²⁰, i.e., is it the prosodic module or the syntactic module that determines the realization of the factative? The answer to this question is essential for the architecture proposed in this thesis.

Evidence for the intonational phrase boundary as a trigger for the factative endoclititic comes from the definite marker *yφ* which usually appears directly behind the noun. However, this sequence can be interrupted by a subordinate clause. In this case, the marker can appear behind the factative clitic ((74), round brackets indicate prosodic phrasing and square brackets indicate syntactic phrasing).

- (74) Owéy ([nú ǎ̀bo ì=vúwóy]_{REL} yφ)_i
 person that hands 3Pl=be dirty.FC DEF
 'The person whose are dirty hands' (modified Kari 2004, 202)

- (75) Mì=món owéy ([nú (baw) e=kótú=n]_{REL} yφ)_i
 1Sg=see=FC person that they 3Pl=call=FC DEF
 'I saw the person who they called.' (modified Kari 2004, 53)

In the above example, the definite marker *yφ* is phrased into the same intonational phrase as the clitic (Kari, p.c. 7-9-2014). Syntactically, however, the definite marker is external to the relative clause, as it belongs to *owéy* 'person'. If the form of the clitic was indeed determined syntactically, then the phrase-final form of the factative should be realized; i.e., endoclisisis in (74) and ↓*un* in (75). Instead, the clitic is realized as if in medial phrase position: It does not emerge between two consonants ((74), cf. (68)) and it is realized as *n* following a vowel and preceding a consonant ((75), cf. 69)). These realizations of the clitic follow from the fact that the clitic indeed is not final if considering not the syntactic, but the prosodic phrasing. Figure 5.10 shows part of the speech signal of example (75).

²⁰Kari (2012) explicitly takes the syntactic clause as the relevant domain for endoclisisis. However, he neither discusses examples as the ones in (74) and (75), nor does he consider prosodic domains in general.

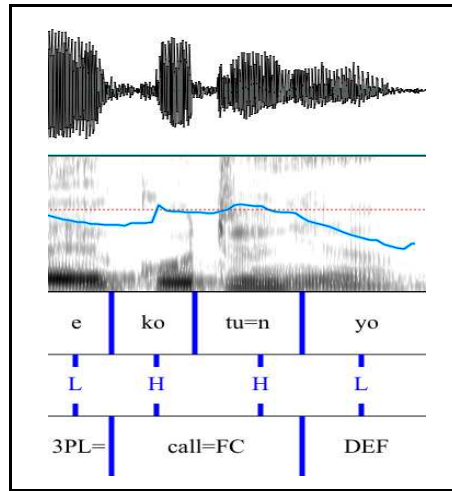


Figure 5.10: Speech signal for $e=kótú=n/REL\ yo)_l$ ‘... (they) called DEF’ ((75)).

The crucial section ($tu=n$) neither consists of two syllables, nor does the pitch show a downstepped high tone. From these facts it can thus be concluded that the relevant domain for endoclisisis in Degema is the intonational phrase (which is almost always isomorphic with the syntactic clause).

5.3.4 Interim summary – paradigm

From the phonological perspective, there are two main factors that determine the shape of the factative clitic: The position of the factative in the intonational phrase and the preceding and (in medial phrase position) the following segment. An overview is given in Table 5.4:

phrase position	phonological environment
medial	$xxV=n\ Cxxx$
medial	$xxV=n\ Vxxx$
medial	$xxC=n\ Vxx$
medial	$xxC\ \emptyset\ Cxxx$
final	$xV^\downarrow VC$
final	$xV=^\downarrow Vn$

Table 5.4: Paradigm of the factative clitic in Degema.

Other phonological processes that apply are vowel harmony between the host and the clitic, metathesis of the clitic into the stem of the host, deletion of segments in various contexts, and high tone downstep between two adjacent high tones. All of these are part of postlexical phonology as it was defined in Chapter 4. The following section provides an analysis of the factative clitic within the syntax-prosody interface proposed in this thesis.

5.4 Degema en(do)clisis: An analysis within LFG

The transfer process of the relevant information from c- to p-structure is two-fold: First, the *transfer of vocabulary* transfers the information stored in the lexical p-form to the preliminary p-diagram (the input to p-structure). Second, the *transfer of structure* conveys the information on syntactic constituents in the shape of prosodic domains to the preliminary p-diagram as well. The combined information is then processed p-structure-internally via the application of prosodic rephrasing and postlexical phonological rules. Figure 5.11 shows the underlying architecture (as it was discussed Section 5.1.3.2) and the (linear) arrangement of the modules with the multidimensional LEXICON as an (independent) look-up instrument.

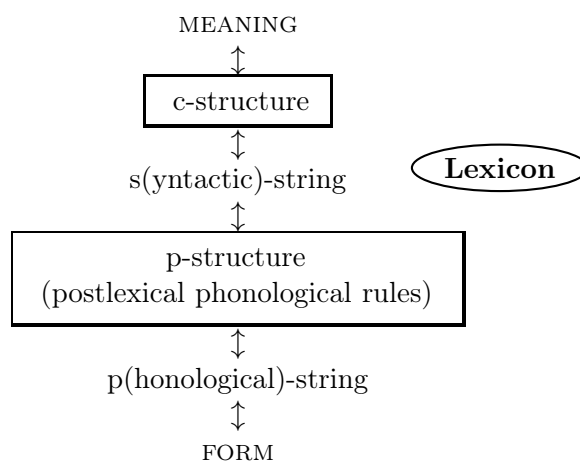


Figure 5.11: Arrangement of modules as assumed in this thesis.

In the following sections these assumptions will be discussed in detail by means of the following example:

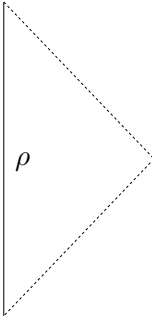
- (76) ómó yo ɔ=só=↓ɔ=1
 child DEF 3Sg=jump₁=FC=jump₂
 ‘The child jumped.’ (Kari, p.c.)

5.4.1 Transfer of vocabulary

The *transfer of vocabulary* refers to the transformation of syntactic string atoms (s-forms) to phonological representations (and vice versa, see Chapter 3). At the heart of this process lies the multi-dimensional lexicon which contains the respective information and serves as a medium for the lexical look-up. In the case of a vocabulary transfer, the syntactically tokenized s-string is divided into s-forms which are matched with the related p-form in the lexical entry. The information gathered in the p-form is then projected syllable by syllable via the relation ρ to the (preliminary) p-diagram. This includes the value of the respective syllable, the lexical tone,

and the metrical frame²¹. With reference to (76), Figure 5.12 shows the transfer of vocabulary between the s-string and the (preliminary) p-diagram. Note that the s-string consists of elements in their syntactic form (e.g., Vn) and that the syntactic linear order shows them *previous* to the application of postlexical phonological modifications.

s-string: ɔmɔ yo ɔ sɔl Vn



s-form	p-form
ɔmɔ N (↑ PRED) = ‘child’	SEGMENTS /ɔ m ɔ/
...	METRICAL FRAME (σσ) _ω
...	LEXICAL TONE HH
sɔl V (↑ PRED) = ‘jump’	SEGMENTS /s ɔ l/
...	METRICAL FRAME (σ) _ω
...	LEXICAL TONE H
Vn CL (↑ STATE) = factative	SEGMENTS /V n/
...	METRICAL FRAME =σ
...	LEXICAL TONE H

PHRASING	(σ	σ) _ω	(σ) _ω	σ=	(σ) _ω	=σ	
...		
L. TONE	H	H	L	L	H	H	
SEGMENTS	/ɔ/	/mɔ/	/yo/	/ɔ/	/sɔl/	/Vn/	
V. INDEX	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	...

Figure 5.12: Transfer of vocabulary: ɔmɔ yo ɔsɔ[↓]l ‘The child jumped’.

After the transfer of vocabulary is completed, nothing within the preliminary p-diagram indicates prosodic phrasing above the level of the prosodic word – up to now, only lexical p-form information has been transferred. In order to realize the factative clitic in its correct form, however, the information on higher prosodic domains has to be available as well. This is achieved via a transfer of c- to p-structure.

5.4.2 Transfer of structure

Under the assumption that syntax influences prosody at least on the higher levels, this thesis follows the proposal by Selkirk (2011) in that every syntactic clause (S/CP) matches an intonational phrase (ι) and every syntactic phrase (XP) matches a phonological phrase (see Chapter 2, Section 2.5.2.4). This means that if there is a CP/S in c-structure, then there will be an intonational phrase wrapping the material formerly contained in the syntactic clause in the p-diagram etc.²² This information is

²¹As discussed in Chapter 4, clitics are indicated by = encoding their prosodic dependency either to the right (proclitics) or to the left (enclitic).

²²This is, in principle, only a translation of inherently syntactic terms which has been criticised by, e.g., Scheer (2011). However, there is no denying syntactic influence on prosodic formation

projected from the respective syntactic nodes via the following annotation pattern:²³

$$\begin{array}{c} S \\ (\natural(T(*)) S_{max} \text{ PHRASING}) =)_l \end{array}$$

The relation \natural is defined as $\rho(\pi^{-1})$ (cf. Bögel 2013, see also Chapter 3, Section 3.5) and directly projects information from c-structure to the p-diagram. This annotation can thus be read in the following way: Consider all the terminal nodes of the current node ($T(*)$) (in that case all the terminal nodes of S). From these nodes take the syllable with the maximum index ($S_{max}=S_6$ in the p-diagram). Add the value $)_l$ (a right intonational phrase boundary) to the attribute PHRASING at this specific index position. The result is shown in Figure 5.13, where the information about the higher prosodic units is added to the already present prosodic units below the word-level.

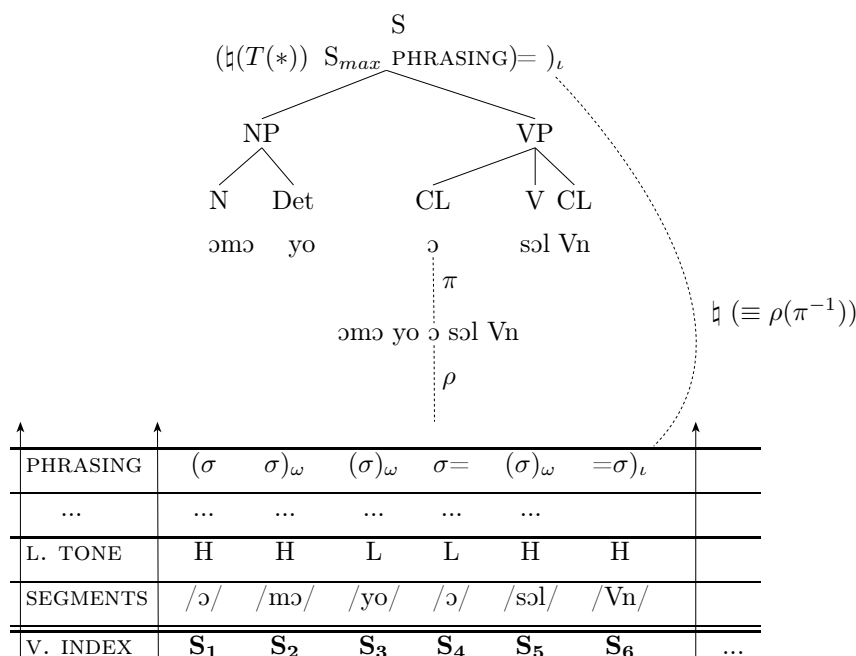


Figure 5.13: Transfer of structure: *ám yo ó sól Vn* ‘The child jumped’.

and as such, the preliminary p-diagram represents the prosodic phrasing as it is determined by syntactic phrasing. It will soon become clear how these ‘mere translations’ are transformed into independently justified prosodic units.

²³Note that the left intonational phrase boarder should be assigned here as well; however, as the left l boundary (as well as the phonological phrase boundaries) is not essential to the analysis an annotation was neglected for reasons of simplification. The respective annotation for the left intonational phrase would be $(\natural(T(*)) S_{min} \text{ PHRASE}) = (, l$, which places the intonational phrase boundary with S_1 .

This preliminary p-diagram, encoding *structure* and *vocabulary* of (76) preserves the syntactic linear order – it is mainly a ‘translation’ of syntactic terms into phonological terms. It allows for reference to segments and their respective features as well as to a preliminary version of prosodic constituency. However, missing is a completion of prosodic phrasing, an explanation of endocclisis, and of other postlexical processes. As will be shown in the following section, these final adjustments are accomplished p-structure-internally via the application of a set of postlexical phonological rules operating on the preliminary version of the p-diagram.

5.4.3 Postlexical phonological processes in Degema

As has been shown in Section 5.3, the Degema factative clitic’s realization depends on several segmental and suprasegmental factors. One relevant factor is the last segment of the host, where the crucial distinction is between vowel and consonant. The same is true for the first segment of the following word (if present). Another factor is the medial vs. final position of the factative clitic in the intonational phrase. This section shows how the relevant postlexical phonological rules operate on the preliminary p-diagram generated in the previous sections via the transfer of vocabulary and the transfer of structure.

As a first step, prosodic phrasing is adjusted (see also Chapter 2). This adjustment always has to precede any other processes because the formation of prosodic domains is vital to the application of most postlexical phonological rules, as they are often blocked or triggered by prosodic constituents. Strictly speaking, the prosodic rephrasing is thus not part of the postlexical phonological rules, but a separate sub-component within p-structure which applies to the preliminary phrasing as it was transferred from lexicon and syntax. This micro-component is subject to different (language-dependent) rules and constraints as they are described in the respective literature (see, e.g., Truckenbrodt (2007) for an overview on relevant factors).

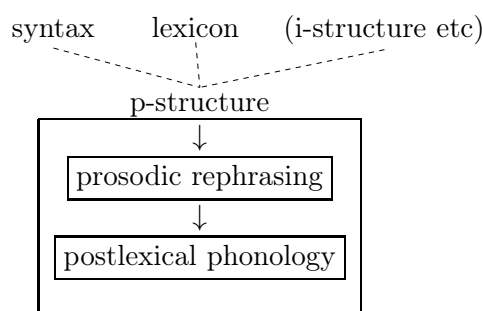


Figure 5.14: The micro-components of p-structure.

Note that these constraints on prosodic phrasing are not only determined by the

modules discussed in this thesis, syntax and lexicon, but also by, e.g., information structural constraints – a research field in itself (Butt and King 1997, Dalrymple and Nikolaeva 2011). In this thesis, processes of prosodic rephrasing will thus be restricted to the phenomena discussed, and will be, for reasons of simplification, collapsed with the rules of postlexical phonology. If the prosodic rephrasing is indeed a separate subcomponent preceding the postlexical phonological rules (strictly ordered subcomponents) or if it is, in part, also determined by the output of these rules (one common component) is left for further research.

It is assumed that the factative clitic forms a prosodic word together with its host. There are two options for this assumption (following also Chapter 4, Section 4.2.2.1, which is based on Selkirk (1995)): a) The clitic can be treated as an *affixal* clitic, in which case it will form a nested prosodic word structure with the host ($((\text{host})_{\omega} \text{clitic})_{\omega}$). Or it can be understood as b) an *internal* clitic forming one prosodic word together with its host $((\text{host clitic})_{\omega})$. As has been shown, the factative clitic often forms the coda of its host or, in the case of endocclisis, is even found within the host. In other phonological environments, it forms a separate syllable or resyllabifies to the following word and the host constitutes an individual prosodic word. Consequently, both clitic types can be argued for. However, to assign two clitic ‘types’ to one clitic would be peculiar and difficult to motivate.

A comparison with other enclitics of Degema that appear in the same position shows that all of them always form at least one separate syllable, and none of them are subject to endocclisis (Kari 2003, 93ff). One possible explanation for this could be that endocclisis only applies to internal clitics and that the factative clitic is indeed an internal clitic and the other enclitics form nested prosodic word constructions. It is for this reason that the factative clitic is treated as an internal clitic in this thesis.²⁴ The following rules realize this assumption and apply the postlexical phonological rules as they were formulated in Section 5.3.²⁵

1. if factative clitic present then incorporate into prosodic word domain of host:

$$\dots)_{\omega} =\sigma \longrightarrow \dots =\sigma)_{\omega}$$
2. if factative in medial intonational phrase position, then realise as n ; delete in context C _ C:

²⁴For a complete analysis of Degema enclitics and the resulting prosodic phrasing, the clitic’s p-form entry in the lexicon would have to be extended with prosodic subcategorization frames (Inkelas 1990). For the factative clitic, the frame would be $[\text{host } _ _]_{\omega}$, while the subcategorization frames for the other enclitics would be $[[\text{host}]_{\omega} _ _]_{\omega}$. Depending on the lexical prosodic subcategorization frames, the postlexical prosodic phrasing would then choose the respective phrasing. However, as this chapter is solely analysing the factative clitic, this aspect is left out in order to simplify the discussion.

²⁵Where $)_{\omega}$ stands for a right prosodic word boundary, $)_i$ for a right intonational phrase boundary, V stands for Vowel, C for Consonant and v for an underspecified vowel.

$$=vn \rightarrow n / \left\{ \begin{array}{l} C_ V \\ V_ C \\ V_ V \end{array} \right\}$$

and $=vn \rightarrow \emptyset / C_ C$

3. if factative clitic in final intonational phrase position, then realise as vn ²⁶:

$$=vn \rightarrow =vn / [\dots] _)_l$$

3.1. if host ends in C, then swap position with C and delete n :

$$Cvn \rightarrow vC / _)_l$$

4. apply vowel harmony²⁷:

$$v \rightarrow V_i / V_i _ C)_l$$

5. apply high tone downstepping if vowels are adjacent:

$$[+H][+H] \rightarrow [+H][\downarrow H]$$

The application of these postlexical phonological processes to the preliminary p-diagram generated in Figure 5.13 generates the final p-diagram for example (76).

↑	↑	↑		↑	↑	↑	
PHRASING	...	$(\sigma)_\omega$	$=\sigma)_l$	\Rightarrow	PHRASING	...	$(\sigma \ \sigma)_\omega)_l$
L. TONE	...	H	H		L. TONE	...	H \downarrow H
SEGMENTS	...	/sɔl/	/Vn/		SEGMENTS	...	[sɔ] [ɔl]
V. INDEX	...	S₅	S₆		V. INDEX	...	S₅ S₆

Figure 5.15: The preliminary (left) and final p-diagram (right) of example (76).

This final p-diagram can be taken to be the combined information provided by syntax, lexicon and postlexical phonological rules to the speech-signal-in-production.²⁸

5.4.4 The complete architecture of the syntax-prosody interface

The following section summarizes the findings and shows the resulting complete architecture of the syntax-prosody interface (Figure 5.16).

²⁶This rule is dispensable, as the underlying form is assumed to be vn . The rule is thus only listed for reasons of ‘completeness’.

²⁷In this case the preceding vowel with all related features is completely copied to the underspecified v – note, however, that the general vowel harmony process in Degema is restricted to the [+/-ATR] feature.

²⁸Note that while the first version of prosodic phrasing definitely can be called ‘diacritic’ (i.e., a mere translation of syntax as claimed by Scheer (2011)), the last version has been subjected to a cascade of postlexical phonological rules. The final p-diagram is thus generated from a combination of syntactic pre-setting and phonological factors and cannot be called ‘diacritic’ anymore.

The architecture proposed in Figure 5.16 shows how postlexical phonology and prosody can be integrated into LFG from c-structure to p-structure. The upper part represents c-structure, where the factative clitic is placed after the verb, or if present, after the object pronoun according to (regular) c-structure rules (VP \rightarrow ... V (Pron) (CL)). The lower part, consisting of the two p-diagrams and the set of postlexical phonological rules, represents p-structure. In accordance with the assumptions made in the introduction about the directionality of each act of language, in the case of *production*, the ‘upper’ p-diagram represents the *input* to the p-structure module, consisting of information from two other grammar components: syntax and lexicon.²⁹ During *comprehension*, the relation is reversed: In that case, the ‘lower’ p-diagram, which is the representation of the speech signal, constitutes the input to p-structure.

The *postlexical phonological rules* are the constraints and rules stored in the p-structure module. They apply to the input, which is the information stored in the preliminary p-diagram (production or comprehension) in a cyclic manner (indicated by \circlearrowright), as the output of one rule often creates the context that triggers another rule. In particular, the creation of prosodic domains has to be completed before the postlexical phonological processes can apply, as they are mostly restricted to a specific domain. But the postlexical phonological rules have to have an internal order as well: In order to apply the high tone downstep, for example, the clitic must first be realized in its full form (Vn) in the final position of an intonational phrase so the vowel hiatus necessary for the downstep to apply is given. For this reason, (postlexical) phonological rules are members of an ordered cascade, which, in principle, can include sets of rules applying in parallel, but which allows for the application of rules to contexts that have been altered by previous rules. Note also that postlexical phonological rules can apply in both directions: In an act of production, where they generate the respective surface form of a string, and in an act of comprehension, where they disentangle a given surface form to allow for a following match-up in the lexicon and a syntactic analysis.

This is of course a simplified description of language production and comprehension, as not all processes are reversible in a 1:1 relation, but there are two major insights. First, the underlying concept of p-structure stays the same: There is an input to which rules and constraints are applied to form an output, and second: It is vital for any linguistic analysis to state the perspective from which an analysis is conducted. The following section discusses this in more detail.

5.4.5 A note on comprehension and production

Architectural assumptions should always consider both: comprehension and production. The above analysis described the interface from the production side. This section will briefly discuss some of the differences that arise if both directions, com-

²⁹The integration of information from other modules is left for further research.

prehension and production, are considered.

The first difference affects the representation of the p-diagram: Depending on comprehension or production, the representation changes slightly: In comprehension, concrete facts from the speech signal itself can be encoded, e.g., the Hertz values of the pitch, but not lexical information (e.g., lexical tone) as this information is ‘not available’ as of yet, but is part of the process that matches strings of speech with lexical entries.³⁰ In the case of a speech signal string that contains the segments /ugo/, there are three available options in the lexicon. On the basis of the tone combinations related to the two syllables of /ugo/ (HL, H[↓]H or LH) in the speech signal, a *lexical tone* combination and with it the correct lexical entry for /ugo/ can be chosen.³¹ In contrast, from the perspective of production, this data from the speech signal is not available, but needs to be ‘produced’. It is retrieved from the lexical p-form and generates, combined with syntactic structure and postlexical phonological rules, the final p-diagram.

Postlexical phonological processes that cause, e.g., the formation of an endoclititic in production are reversible because the corresponding rule can be applied backwards, separating the two lexical items out of an incoming speech signal so they can be matched against their respective lexical entries. The *transfer of vocabulary*, mediated via the multi-dimensional lexicon, is also similar in production and comprehension.

In contrast, the *transfer of structure* differs between production and comprehension. From the perspective of production, the transfer of structure gives a first indication of prosodic phrasing, which is retained, if syntax and prosody happen to be isomorphic. However, in cases like (74), repeated in (77), where the two components are non-isomorphic, prosodic phrasing is adjusted as part of the postlexical phonological processes (Table 5.5).

- (77) Owéy ([nú ábo i=vúwóy] yo)
 person that hands 3Pl=be dirty.FC DEF
 ‘The person whose hands are dirty.’ (modified Kari 2004, 202)

1.	<i>Transfer of structure:</i> Every CP/S projects an int. phrase	[Owéy ([nú ábo i=vúwóy] _{CP}) _l yo] _{NP}
2.	<i>Prosodic Rephrasing:</i> Adjustment of int. phrase	[Owéy ([nú ábo i=vúwóy] _{CP} yo] _{NP}) _l

Table 5.5: Structure transfer and prosodic rephrasing in cases of non-isomorphism.

During comprehension, on the other hand, the corresponding speech signal phrases *yó* together with the preceding intonational phrase of the embedded clause without any reference to syntactic boundaries: ... *nú ábo i=vúwóy yó*)_l

³⁰See, e.g., McQueen (2005) for an overview on this process.

³¹See Section 5.2 for more details on lexical tone and the word *ugo*.

The information that a CP boundary is present between *yφ* and *vúwóy* is only available at a ‘later stage’, i.e., when the respective string is parsed syntactically.

As a consequence, it can be assumed that prosodic adjustments, applied during production, are not ‘reversible’ during comprehension. A further conclusion would then be that prosodic phrasing is not necessarily vital for the parsing of syntactic constituents – i.e., a non-rephrased intonational phrase boundary after the CP in example (77) will not result in a failure of syntactic phrasing, albeit prosodic phrasing *can* reflect syntactic phrasing and aid in the disambiguation of syntactic ambiguities (see Chapter 3).

5.5 Summary and conclusion

This chapter demonstrated how the linguistic issues caused by endocclisis (namely the violation of lexical integrity and modularity) can be resolved by assuming a non-parallelity between the p-string and the s-string. The underlying assumption is that the s-string represents syntactic linear order as it was determined by c- and f-structure, but that the p-string represents the linear order of the elements as determined by prosodic constraints. While p- and s-string are mostly parallel, the p-string can differ from the s-string in cases where prosodic constraints require the reordering of specific elements.³²

This reordering process is part of p-structure where prosodic and phonological constraints are applied to the input provided by syntax and lexicon (and other structures outside of the scope of this dissertation). The transfer process proposed here is thus two-fold: a) the *transfer of structure* transfers phrasal information from c-structure to p-structure, and b) the *transfer of vocabulary* ‘translates’ the syntactic elements of the s-string into the respective p-form and transfers all the information associated with that p-form. Combined, these transfer processes form the c-structure–p-structure interface (or more commonly phrased: the syntax–prosody interface).

Within p-structure, the preliminary prosodic structure is rephrased according to language-internal constraints, and a set of postlexical phonological rules is applied. The output, the ‘final p-diagram’, is the contribution of syntax, lexicon and postlexical phonology to the speech signal; the resulting linear order of the elements is the surface (p-)string. These architectural assumptions allow for an analysis of endocclisis independently of syntax; thus, the modularity of the different components is preserved and the concept of lexical integrity is kept intact.

It has also been shown that it is vital for the analysis to take a specific directional perspective: from meaning to form or vice versa. Each direction should be considered

³²It could be that ‘special’ clitics are the only elements that can cause a misalignment of prosodic and syntactic structure, although processes like heavy NP shift and the like should be considered here as well.

during the analysis in order to uncover wrong conclusions that are not identifiable as such from a monodirectional perspective.

Returning to the initial question as to how the s- and p-string should be defined and how they relate to each other, it can be concluded that s-string and p-string are parallel most of the time, but they do not necessarily have to be. In cases of endoclisis and other postlexically reordered elements, the linear order of the p-string may change. The s-string is thus defined as the *linear order of elements as they would be syntactically analysed* while the p-string represents *the linear order of the elements as they would be pronounced* in the final p-diagram. With the architectural assumptions proposed in this chapter, this difference in linear order can be easily explained and forms a stable construction for the analysis of complex phenomena at the syntax-prosody interface.

The following chapter will discuss a further case where the linear of the surface string is different from the linear order given in c-structure: Pashto second position en(do)clisis.

Chapter 6

Second Position (endo)clisis: Pashto

6.1 Introduction

Pashto, an eastern Iranian language spoken by approximately 50 Million speakers in Afghanistan and Pakistan has an interesting kind of forms from the perspective of the syntax–prosody interface: second position en(do)clisis. The present chapter introduces Pashto second position (henceforth 2P) clitics and analyses them in the model proposed in the previous chapters. Pashto 2P clitics have been discussed in numerous articles and dissertations, mostly with a focus on a specific phenomenon that results as part of the 2P clitic placement: endocclisis (See also Chapter 5). However, most accounts are not able to completely cover the data, which is also accounted for by the very complex constraints on 2P clitics in Pashto, as the elements that occupy the first position can be anything from a syntactic constituent of considerable size down to a part of a verb. The only common criterion seems to be for the host to carry some accent, but there are several confusing examples. As a result, most accounts have either focussed on the prosodic *or* the syntactic side. However, none has taken the crucial factors for the analysis presented in this chapter into account: preverbal (pro)clitics and postlexical phonological processes.

The solution presented here is new in that it clearly distinguishes two separate processes of 2P clitic placement: One determined solely by syntactic terms, and one ‘last-resort’ postlexical prosodic clitic placement. The separation of these two processes is accomplished by a close analysis of the syntactic structure in Pashto, in particular of the verbal complex and the associated preverbal clitics, in relation to postlexical phonological processes. The result is a straightforward solution to formerly confusing data.

This chapter will first give a brief introduction to second position clitics and prosodic inversion, before providing an overview of the Pashto data in Section 6.2, followed by a description of some of the most influential previous LFG and non-LFG

accounts of the data in Section 6.3. These sections are followed by a close inspection of Pashto syntax with a focus on preverbal proclitics and a description of several phonological processes that give insight into the prosodic phrasing of especially the verbal complex. The findings result in a new proposal which allows for a straightforward solution of Pashto second position clitics. This solution is implemented in the last Section 6.5.

6.1.1 Second position clitics

This section provides a brief introduction to the phenomenon of second position clitics.¹ Second position clitics were first described in a larger study in the work of Jacob Wackernagel (1892) who notes that enclitic elements in Ancient Greek have a strong preference for the second position after the ‘first word’ of the sentence. In his work, he also takes note of the promiscuous attachment, the ability to break up (syntactic) phrases, and the prosodically deficient status of second position clitics (cf. Spencer and Luís 2012, 41).

Subsequent work has extended the notion of the ‘first word’: “the two most common cases seem to be either the first phonological word or the first syntactic daughter of the constituent within which the clitic is second” (Halpern 1995, 15). In the introduction to a collection on different second position phenomena, Halpern defines a second position clitic as “an unstressed, closed-class element which, unlike nonclitic elements of similar function, must appear second in a clause” (Halpern 1996, x.). Although this seems to be a very clear definition, Halpern lists several issues concerning the definition of second position clitics: a) the clitics often do not appear in the linear second, but in the third or even a later position; b) the definition of ‘second’ varies greatly between languages (and even within languages, as will be discussed in this chapter). It might range from a syntactic constituent to a small prosodic unit. And c), some hosts that are seemingly perfect for an enclitic to attach to, are unsuitable for no apparent reason.

Next to these issues concerning the TYPE of the host, there are also issues concerning the derivation of the linear order. Some researchers account for the positioning in purely syntactic terms, others analyse the phenomena from a prosodic perspective, and a third group mixes the two approaches. It is thus no wonder that “The precise definition of ‘second position’ is [...] an exceedingly complex and contentious matter” (Franks and King 2000, 28). Franks and King provide an overview of Slavic clitics and show that for most second position clitics, syntax is an important factor, but that the syntactic arrangement might be manipulated on grounds of prosodic

¹Note that the research field on second position clitics is very wide and that this brief introduction by no means aims (and claims) to give a full overview: Instead, centrally relevant notions are introduced. For some work on second position clitics in LFG, see King (2005) and references therein, Bögel et al. (2010) for an account of Serbian-Bosnian-Croatian 2P clitics, and Lowe (2011) and Lowe (forthcoming), both with reference to OT. The latter will be discussed in more detail in Section 6.3.2.

constraints. This is also the essential result of the analysis of Pashto 2P clitics in this chapter.

Both types of special clitics, endoclitics and second position clitics, are problematic from the perspective of the string as they can alter the default linear order of the elements on grounds of prosodic/phonological constraints.² As has been discussed in Chapter 5, this problem can be resolved if the s-string and p-string are seen as parallel most of the time, and if the linear order can vary as a consequence to the intervention of prosodic constraints.

6.1.2 Prosodic inversion: Halpern (1995)

In his thesis, Halpern (1995) develops an account of second position clitic placement. Halpern assumes that clitics are associated with a particular syntactic position, but that “unlike other words the position of a clitic in the surface string of a sentence may diverge from what would be expected based on its syntactic position” (Halpern 1995, 17). The reason for this behavior is that clitics are different from other elements in that they lack stress and thus need to be combined with a host *postlexically* (Halpern 1995, 14).³ This attachment to the host is accounted for by a process which Halpern terms *Prosodic Inversion* (PI), where a syntactically initial (en)clitic linearly switches position with the next available element to its right, its ‘host’. This process is illustrated in Figure 6.1 (where the trace indication t_i is *not* a syntactic process, but describes a postlexical prosodic inversion).

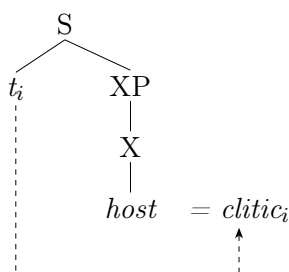


Figure 6.1: Postlexical prosodic inversion of a syntactically initial clitic.

This analysis is convenient for a theory like LFG, because it allows for functional information to be gathered before the (en)clitic is moved postlexically into its prosodically determined position (see e.g. Austin and Bresnan 1996, Nordlinger 1998, Bögel et al. 2010). In the abstract example in Figure 6.1, this means that the functional information projected from the enclitic is part of the clausal scope during syntactic analysis (t_i), while the postlexical prosodic placement might place the clitic in

²This is not necessarily true for *all* occurrences of special clitics, as some of them seem to appear in a specific position for, e.g., syntactic reasons, which is to some extent also true for Pashto 2P clitics. In this case, there is no mismatch in linearity.

³However, see Plank (2005) and references therein for an alternative approach.

a position where the functional information of the enclitic is no longer available to the top level of f-structure, but would be restricted to the f-structure projected by XP. However, with a differentiation between syntactic analysis and postlexical prosodic placement, this is not problematic, as the enclitic's functional information was analysed before its position in linear order was determined by postlexical prosodic constraints.

As to the nature of the host, Halpern describes several languages that feature second position clitics. He establishes that for some it is clearly the first (prosodic) word, and for others the first syntactic daughter. Some languages seem to feature both⁴ and allow for a relative free variation, while others also feature both, but with very clear restrictions, among them Pashto.

As Halpern views the host in second position as always prosodically determined, the question why the clitics attach to the first word in some cases and to the first constituent in other cases is a remaining issue in his approach. One strategy proposed is the claim that the host is part of a domain which is inaccessible to the clitic; thus, the clitic is forced to attach to a larger prosodic domain (Halpern 1995, 72). However, the claim that host-constituents with considerable size as they appear in Pashto can be prosodically paralleled to host-constituents that consist of one word or less is questionable. As will become clear below, the patterns can be fully explained by assuming a *syntactic* second position that corresponds to Halpern's *syntactic daughter* and a *prosodic* second position placement that corresponds to Halpern's *prosodic word*.

As a conclusion, the approach presented here does not apply prosodic inversion to the syntactic placement; instead, syntactic second position is determined syntactically. However, with the *prosodic* placement of Pashto 2P clitics as discussed below in Section 6.4.4, prosodic inversion is exactly the process that yields the correct results.

6.2 Pashto second position: An overview

The group of 2P clitics discussed in this thesis involves personal pronouns (as they appear in ergative, accusative and genitive constructions), modals and adverbials all of which are listed in the following table.⁵

⁴For example 2Ps in SCB, see also Bögel et al. (2010) and references therein.

⁵The data, its notation and gloss is mostly taken from Tegey's (1977) dissertation on Pashto clitics. These examples have been rediscussed in the subsequent literature and complemented by further examples. It has also been noted that Pashto has several dialects, some of which do not allow the formation of endocclisis, but an exact account of the regional distribution awaits further research. In order to clearly indicate clitics in these examples, the equality sign (=) was added to denote the direction of the prosodic deficiency (enclisis or proclisis). Furthermore, in order to distinguish 2P clitics from other clitic elements in the language, 2P clitics are written in bold in the following examples.

Weak Pronoun	Num.&Pers.	Modal	Translation	Adverbial	Translation
me	1. Sg	ba	will, should	xo	really
de	2. Sg	de	should, let	no	then
ye	3. Sg				
am / mo	1. Pl				
am / mo	2. Pl				
ye	3. Pl				

Table 6.1: Pashto second position clitics (Tegey 1977, 81).

All of these enclitics show identical behavior with regard to their appearance in second position. If more than two enclitics cooccur, they are placed in a fixed template (Tegey 1977, 191).

(78) 1 2 3 4 5 6 7 8
 xo ba am am/mo me de ye no

While Roberts (2000), whose work is couched within a derivational framework, claims that this order can be derived syntactically and there is consequently no need for a template, Hock (1996) convincingly argues for the opposite, emphasising the fact that a) syntactically related clitics are spread over the template (e.g., modals: position 2 and 6, adverbials: 1 and 8) and b) in position 6, the modal/pronoun *de* can only be realised once: An appearance of identical forms (i.e., *dede*) is prohibited. This suggests a cluster-internal clitic placement determined by phonological constraints (see Hock (1996, 212) for further discussion).⁶

Note also that all of the 2P clitics in Table 6.1 are expected to have functional scope over the complete clause within which they are placed. This suggests that the clitics are immediate daughters of S, a fact that will become important in the discussion in Section 6.4.

The following sections will provide an overview on the distribution of 2P clitics in Pashto, first describing syntactic constraints, before discussing prosodic factors and a resulting construction thereof that has been in the focus of research for decades: endocclisis.

6.2.1 Syntactic constraints

At first glance, the 2P clitics seem to be simply placed after the first word of a sentence. In (79), the pronoun clitic *ye* (3Sg, ‘he’) follows the noun *angur* ‘grapes’ while the modal clitic *ba* ‘maybe’ is placed after the adjective *naroġa* ‘sick’ in example (80).

⁶In the following sections, the 2P clitic cluster will be united under the syntactic node CCL. CCL stands for ‘clitic cluster’ and was proposed by Bögel et al. (2010) as a collective node for 2P clitics that appear in a templatic order. See Section 6.4.3 for more information.

(79) angur =**ye** rəwɾə
 grapes he brought
 ‘He brought grapes.’ (Tegey 1977, 138)

(80) nəroğa =**ba** wi
 sick maybe is
 ‘Maybe he is sick.’ (Tegey 1977, 84)

However, the host of the enclitic is not necessarily the first word of the sentence. Example (81a) shows that the element serving as a host for the enclitic can be a syntactic constituent, in this case a coordinated noun phrase. This coordination may not be interrupted ((81b)).

(81) a. [xuʂal aw patang] =**ba** =**ye** dər= ta= rəwɾi
 Koshal and Patang will it you to bring
 ‘Koshal and Patang will bring it to you.’ (Tegey 1977, 84)
 b. *xuʂal =**ba** =**ye** aw patang dər= ta= rəwɾi

The same is true for postpositional phrases, where the enclitic is not allowed in between the postposition and its argument.

(82) [layla na] =**de** axistə (*layla =**de** na axistə)
 Layla from you buy
 ‘You were buying it from Layla.’ (Tegey 1977, 114)

The initial constituent can be of a considerable size as well, without the enclitic interrupting.

(83) [ağə ʂəl kaləna xəysta peğla aw loy təgay ələk] =**me** nən bya wəlida
 that 20- year pretty girl and big thirsty boy I today again saw
 ‘I saw that pretty 20-year old girl and the big thirsty boy again today.’

Such constructions are in principle unrestricted and infinitely extendable; as long as the requirement of a ‘first syntactic daughter’ (i.e., the first constituent that is an immediate daughter to S) is met, the 2P clitics are always placed after that constituent. Note also that with examples like (83), it is difficult to argue for a common prosodic domain, e.g., the phonological phrase, as, especially with coordinations of NPs that contain many modifiers (e.g., as in (83)), one would expect the constituent to contain at least two phonological phrases.

In constructions involving subordinate clauses, the 2P clitic may not appear outside of the clause in which it functionally originates, but instead is inserted after the first element therein.⁷

⁷See Pate (2012) for a very good overview on subordinate clause constructions in Pashto, also in relation to clitics.

- (84) hayə səʕei [tʃe [pʊʂtənə =**mi** dzine wəkʕə]] zmā plār dəi
 that man COMP question I from PERF.do my father be
 ‘That man that I asked is my father.’ (Pate 2012, 64, modified)

This is also true for the coordinated sentences in (85), where the two clitics are part of their individual clauses, occupying the second position respectively.

- (85) [tor =**me** wəlidə] magar [spin =**me** wə nə lidə]
 Tor I saw but Spin I PERF not saw
 ‘I saw Tor, but I didn’t see Spin.’ (Tegey 1977, 127)

Up to this point, the constraints responsible for the positioning of the 2P clitics can be explained syntactically: The clitics are placed after the first syntactic constituent in the clause. However, as the following section shows, clitics are also positioned according to prosodic constraints.

6.2.2 Prosodic constraints

Besides the syntactic requirements, prosody plays a crucial part as well, as can be seen in (86) and (87), where the 2P clitic is placed after the first element bearing lexical stress.

- (86) rɑ= ta= te= rɑ= ʈolawəl =**de**
 me for from_it here collect you
 ‘You were collecting them for me from it (and bringing them) here.’
 (Tegey 1977, 119)

- (87) rɑ= ta= pe= gɑndə́ =**de**
 me for by_him sew you
 ‘You were having him sew it for me.’ (Tegey 1977, 118)

The elements preceding the verb belong to another group of Pashto clitics, consisting of combinations of oblique pronouns and postpositions (Tegey’s “Type II clitics”), which are usually placed in front of the verb (see Section 6.4.2). These clitics are prosodically unstressed material which forces the 2P clitics to appear after the first stressed element to its right. In the above constructions, the first element of the sentence bearing stress is the verb, which is thus the only possible host for the prosodically deficient 2P clitics even though Pashto is a fairly rigid verb-final language.

However, if the preverbal postposition is stressed it can function as a host for 2P clitics as well ((88a)). If it is destressed, on the other hand ((88b)), the 2P clitic selects the next stressed element.

- (88) a. $r\alpha$ sará =de wi b. $r\alpha$ sara wí =de
 me with let be me with be let
 ‘Let it be with me.’ ‘Let it *be* with me.’
- (Tegey 1977, 121)

The resulting hypothesis that the 2P clitic attaches to the first accent-bearing element of the sentence is also confirmed when it comes to endoclitics. In the perfective version of example (86), the 2P clitic is inserted *into* the verb following the part of the verb that bears the main accent.

- (89) $r\alpha=$ ta= te= $r\alpha=$ [tól =de kɾəl]_V
 me for from_it here collect₁ you collect₂
 ‘You collected them for me from it (and brought them) here.’
- (Tegey 1977, 119)

This type of second position has been in the focus of substantial amounts of earlier literature. The following section will thus give a close account of the data related specifically to endocclisis, before resuming the discussion on 2P clitics on a more general basis.

6.2.3 Pashto endoclitics

Pashto is an argument-dropping language;⁸ thus, sentences can consist of only a verb and a 2P clitic. The endoclitics mainly appear in these short sentences in the context of a stress alternation that is determined by a difference in aspect.

- (90) a. **perfective:** b. **imperfective:**
 ták =me wəhə tək wəhə =me
 shake₁ I shake₂ shake I
 ‘I shook it.’ ‘I was shaking it.’
- (Tegey 1977, 92)

In Pashto, the *perfective* aspect of the verb is accompanied by a verb-internal stress shift placing the main stress on the first foot of the verb, while the verb in the *imperfective* aspect carries the main stress on the last foot of the verb. With regard to the stress shift, Pashto verbs fall roughly into three classes, depending on their word-internal structure. Following Tegey’s classification, their internal characteristics and their behavior concerning the placement of 2P en(do)clitics will be introduced in the following sections.

⁸More on the syntactic structure of Pashto in Section 6.4.1.

6.2.3.1 Monomorphemic verbs: Class I

Class I verbs are monomorphemic. In the imperfective, these verbs bear stress on the last foot (the suffix); the 2P clitic is placed after the verb ((91a)). In the perfective aspect however, the perfective prefix *wə-*, which receives the main stress, is added to class I verbs.⁹ In this case, the 2P clitic occurs after the prefix and in front of the stem ((91b)).

- | | |
|-----------------------------|-------------------------------------|
| (91) a. imperfective | b. perfective |
| təxnawəla =me | wə =me təxnawəla (*wətəxnawəla =me) |
| tickle I | PERF I tickle |
| ‘I was tickling (her).’ | ‘I tickled (her).’ |

(Tegey 1977, 86)

In contrast to class I verbs, class II and III form the perfective by means of a stress shift from the last to the first part of the verb without adding a perfective prefix. The verbs of both classes are bimorphemic in the greater sense, but have certain distinct characteristics.

6.2.3.2 Bimorphemic verbs: Class II

Most class II verbs consist of a derivational prefix and a root. In the imperfective aspect, the stress is on the (pen)ultimate syllable of the verb and the 2P clitic is placed after the whole verb ((92a)). The perfective is formed via a stress shift to the derivational prefix of the verb. The 2P clitic is then placed after the prefix as illustrated in example (92b).

- | | |
|-----------------------------|-----------------------------|
| (92) a. imperfective | b. perfective |
| ʔelwəhə =me | ʔél =me wəhə (*télwəhə =me) |
| push I | PREF I push |
| ‘I pushed (it).’ | ‘I was pushing (it).’ |

(Tegey 1977, 92)

While the data presented so far is not ‘endoclititic’ as it was defined in Chapter 5, Section 5.1.1, but could be described as cases of mesoclysis,¹⁰ there are also verbs in Pashto that are not clearly separable into a prefix and the stem, i.e., there is a group of verbs within class II which do not contain any identifiable derivational prefix and where the second part of the verb (e.g., *lodə* below) also does not constitute a separate morpheme (Tegey 1977, 93/94).

⁹Tegey’s prefix *wə* is called a ‘stressed proclitic’ by Roberts (2000, 24) and an ‘auxiliary’ by Pate (2012). As neither provides evidence and there furthermore is no evidence that suggests that *wə* is *not* a prefix, the account introduced here follows Tegey’s initial classification. The data presented in this chapter concerning the perfective prefix *wə* will support this initial assumption.

¹⁰That is, the insertion between an affix and a stem, see Chapter 5, Section 5.1.1.

- (93) a. **imperfective** b. **perfective**
 báyłódə =me báy =me lodə
 lose I lose₁ I lose₂
 ‘I was losing (it).’ ‘I lost (it).’

(Tegey 1977, 93, modified)

In example (93), the element after which the 2P clitic is placed (in the above example *báy*) does not constitute a morpheme with a separate meaning. It is therefore rather difficult to argue in favor of a prefixal or even clitic status of *báy* if the morpheme is not identifiable as such and furthermore holds a unique position within the language as it cannot be found in any other word. That is, for this particular group of verbs, real endocclisis exists.

The following figure shows the corresponding speech signal for (93a) and (93b), where the imperfective is stressed on the second part of the verb with the 2P clitic following, while the perfective clearly only carries stress on the first partial. The 2P clitic and any leftover material of the perfective verb is destressed.¹¹

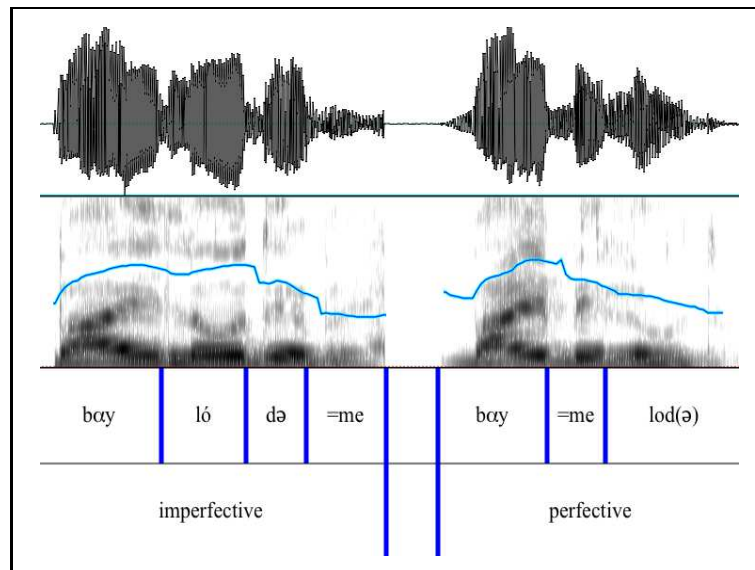


Figure 6.2: Speech signal for (93a) and (93b): *báyłódə+me*.

¹¹The speech signal is part of a first set of recordings of Nafees Ur Rehman at the University of Konstanz in spring 2015.

6.2.3.3 Complex predicates: Class III

Class III verbs are complex predicates consisting of an adjective, adverb or noun and a light verb¹² and form the largest group of verbs in Pashto. Their behavior with respect to 2P clitics is the same as with the class II verbs in that the aspect is expressed by a stress shift from the light verb to the lexical content word.

- (94) a. **imperfective**
 ʈolawəl =de
 collect you
 ‘You were collecting (them).’
- b. **perfective**
 ʈól =de kɾəl
 collect₁ you collect₂
 ‘You collected (them).’
- (Tegey 1977, 119, modified)

As the single elements constituting the complex predicate can be clearly distinguished, the 2P clitics in that construction have rather enclitic than endoclititic status. However, note that the complex predicates in the perfective and the imperfective have different phonological characteristics that contribute to the overall determination of the placement of 2P clitics. This will be discussed in Section 6.4.5.2.

6.2.3.4 A hybrid: The a-initial verbs

Within class I, there is small group of verbs that can have alternating stress in the imperfective, but form the perfective with the perfective prefix of class I (*wə*). Within this group, there are verbs that begin with consonants, which do not show any special behavior in the imperfective: even if the stress is on the first part of the verb, the 2P clitic is placed after the verb.

However, there is a small number of verbs in this group with an initial vowel /a/ which show a very distinct behavior with regard to the alternating stress shift in the imperfective. If the stress falls on the second part, the 2P clitic is placed after the verb ((95a)). If it falls on the initial /a/ however, the clitic is placed directly after the vowel as in (95b). The a-initial verbs could thus be seen as an indicator for the hypothesis that it is not the aspectual feature that is responsible for the 2P clitic placement, but rather the position of stress.

- (95) a. **imperfective:**
 aɣustá =me
 wear I
 ‘I was wearing it.’
- b. **imperfective:**
 á =me ɣustə
 wear₁ I wear₂
 ‘I was wearing it.’
- (Tegey 1977, 89)

¹²Tegey names these verbs auxiliaries. The notion *light verb* comes from the literature on complex predicates and is based on the fact that the ‘main predicational content’ is provided by the noun (see Mohanan (1990) and Butt et al. (2012) and references therein).

It has been argued that /a/ is a separate prefix/clitic from a diachronic perspective (Kaisse 1981, Anderson 2005), but this cannot be confirmed for all *a*-verbs.¹³ Furthermore, synchronically, the initial /a/ does not have a recognizable prefix/morpheme-function, as Tegey explicitly states in his thesis (Tegey (1977, 89), see also Kenstowicz and Kisseberth (1977)). The same can be said of the remainder of each form – *ǰust* and all other “remaining” roots are not identifiable as separate morphemes. Hence, additionally to the group of class II verbs where the 2P clitic is inserted after a morphologically unidentifiable item (as in example (93)), there is another group of verbs where real endocclisis occurs.

A-initial verbs are also interesting with respect to their interplay with the phonological process of vowel coalescence as it also provides evidence that the 2P clitic is inserted into the verb postlexically. As has been mentioned before, the a-initial verbs take the perfective prefix *wə-* in the case of a perfective construction. In contrast to the consonant-initial verbs, however, perfective a-verbs display *vowel coalescence*. In example (96a), the adjacency of the perfective prefix /wə/ and the initial /a/ results in a fusion: /wə+/a/ → [wɑ]. In the event of 2P clitic insertion after the perfective prefix, the fused vowel is still present ((96b)).

(96) a. **Vowel coalescence — without clitic:**

tə =ye wɑxla (*wə axla)

you it PERF.buy

‘You buy it.’

(Tegey 1977, 149)

b. **Vowel coalescence — with clitic:**

wɑ =ye xla

PERF.buy₁ it buy₂

‘Buy it.’

(Tegey 1977, 163)

On the basis of examples like (96b) Tegey concludes that 2P clitic placement takes place after the process of vowel coalescence. Table 6.2 illustrates how the perfective prefix /wə/ and the initial /a/ are subject to vowel coalescence, before the 2P clitic is then moved to its position after the first stressed element.

data	process
wə́ axla ... ye	
wá xla ... ye	vowel coalescence
wá ye xla	2P clitic placement

Table 6.2: Tegey’s approach: vowel coalescence before 2P clitic placement.

This particular process has been problematic for transformational accounts of Pashto endocclisis, as the seemingly syntactic movement of 2P clitics takes place *after* the

¹³The verbs where the *a*-vowel cannot be identified as a prefix diachronically are claimed to have been reanalysed (Kaisse 1981).

phonological vowel coalescence (see Kaisse (1985), Section 6.3.1). Furthermore, as mentioned above, several accounts have discussed the status of initial /a/ as a prefix/clitic, because endocclisis as such is a difficult phenomenon to explain in any framework, especially so in frameworks that adhere to the principle of lexical integrity.¹⁴

6.3 Previous accounts

The previous sections gave a brief overview of the data on Pashto 2P clitics. The requirements for the positioning seem to be drawn from (at least) two modules: the (postlexical) phonological module and the syntactic module. The resulting constraints have led to a lot of confusion: On the one hand, 2P clitics seem to follow the first syntactic constituent which they cannot interrupt. On the other hand, 2P clitics seem to be able to move *into* the verb. On the one hand, 2P clitics clearly require for the host to be stressed. On the other hand, however, it is very difficult to unite all shapes of the ‘first constituent’ under one prosodic domain, as it can consist of a partial word host or a complex coordinate structure with several phonological phrases.

These seemingly conflicting constraints on the positioning of 2P clitics in Pashto are reflected in previous accounts of the phenomenon, which usually focus on one module with only scarce reference to the cases that are determined by the other module. The following section discusses some of the proposals.

6.3.1 Previous accounts – outside of LFG

Kaisse (1981, 1985) focusses on one particular aspect of Tegey’s analysis, namely the vowel coalescence occurring with a-initial verbs. Tegey argued that vowel coalescence occurs before the 2P clitic is positioned within the word (Section 6.2.3.4). Working within a derivational framework and supporting a grammar architecture roughly corresponding to the T-model (see Chapter 2, Section 2.4.2), Kaisse assumes that no phonological rule applies before a syntactic rule. Consequently, Tegey’s analysis is problematic, because a phonological rule (=vowel coalescence) operates before a syntactic rule (=the placement of the clitic within the word). In order to resolve this tension, Kaisse argues that the a-initial verbs consist of a stem and a prefix /a/. She rejects an earlier claim that /a/ does not constitute a separate prefix on the grounds that a) meaningfulness “cannot be a necessary criterion for morphem-hood” (Kaisse 1985, 139) and that b) other verbs of class II can also be separated by a 2P clitic even though there is no discernable prefix. The last argument in particular is a circular one, as the argumentation could also be that /a/ is *not* a prefix, *because*

¹⁴Interestingly, the other group of class II verbs that do not consist of a discernable prefix and thus seem to be real cases of endocclisis as well have been largely ignored in the literature.

there are other verbs where the single parts cannot be identified as distinct elements either.

Under the assumption that /a/ is a prefix, Kaisse assumes that stress is assigned to that prefix as part of the morphological interpretation.¹⁵ Following this stress assignment, the 2P clitic is syntactically moved to the position after that prefix. Vowel coalescence is then part of a (postlexical) phonological rule given in (97).

$$(97) [\text{ə}]_{particle} + [\text{a}, \text{a}]_{verb} \rightarrow [\text{a}]^{16} \quad (\text{Kaisse 1981, 202})$$

Table 6.3 summarizes this process.

surface form	process
wə a xla ... ye	
wə á xla ... ye	Perfective Stress Assignment
wə á=ye xla	Clitic Placement
wá=ye xla	Vowel Coalescence

Table 6.3: The generation of vowel coalescence according to Kaisse (1985, 142).

Kaisse thus assumes an architecture like the following:

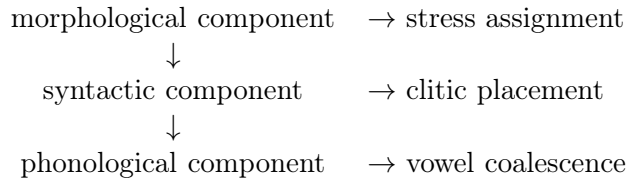


Figure 6.3: Kaisse’s architectural assumption.

The assumption that 2P clitic placement can only be a syntactic process is obsolete; it can be seen as established that many clitics are placed according to prosodic constraints. Furthermore, Kaisse claims that Pashto displays no real endocclisis by arguing that all verbs allowing for endocclisis can be described as bimorphemic, i.e., as containing a prefix after which the 2P clitic attaches. However, as discussed above, this is not possible as there are several verbs where there are no discernable morphemes, neither synchronically nor diachronically.

¹⁵Thereby ignoring the fact that wə- receives main stress in all other cases.

¹⁶Note that the phonological rule in (97) relies on word category information as “the rule cannot be written in purely phonological terms” (Kaisse 1985, 137). In her approach to phonology, Kaisse relies directly on morphosyntactic information (e.g., word categories) to restrict the rule on vowel coalescence to this particular context and to prohibit it from applying elsewhere, e.g., between two words. The analysis in purely phonological terms (i.e., in modular terms, where phonological rules only refer to phonological information) would nowadays rely on a specific prosodic domain, within which the rule applies (e.g., the prosodic word).

Although working within a transformational framework as well, the problem of a-initial words, vowel coalescence and 2P clitic placement is unproblematic for **Roberts (2000)**. He focusses on a syntactic account that views the prosodic placement as a last resort, thus allowing for 2P clitics to move on the basis of prosodic constraints after vowel coalescence has applied — an underlying assumption which is shared by this thesis. Roberts (2000) takes the second position clitics to be agreement morphemes¹⁷ that are not moved in the syntax and describes the perfective prefix *wə-* and the negative marker *nə* as proclitics without analysing the difference between both elements (see below Section 6.4.4.2). The main reason for this classification of *wə-* is the possibility of 2P clitics to be placed between the markers and the verb; thus Roberts evades the notion of ‘endoclisism’ by defining all elements that precede the 2P clitics to be clitics as well – a circular argumentation, because it could be argued that *wə-* is a prefix, because 2P clitics *can be* endoclitics.

For the verbs that provide only a stress shift as an indication of aspectual change, Roberts develops a purely syntactic account in that he declares the perfective aspect as a strong feature that allows for the verb to be moved to ASP, while the imperfective is a weak aspect feature, whose movement is ‘delayed’ until after spell-out. Figure 6.4 illustrates the approach with an example of a class III construction (translated into English).

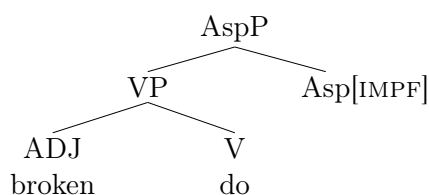
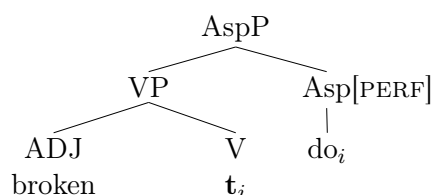
a. **imperfective:**b. **perfective:**

Figure 6.4: The imperfective and the perfective aspect (Roberts 2000, 51, modified).

At PF, the perfective structure in Figure 6.4 (b) is translated into two prosodic words that are then intervened by the 2P clitic, while the imperfective construction in (a) is only represented by one prosodic word; consequently, the 2P clitic is attached after the complete construction.

In Figure 6.5, the same method is applied to the stress-shifting verb *kenastəl* ‘to sit down’ for the following example in the perfective aspect:

- (98) *šagerdan ke nə nəstəl*
 students sat_down₁ not sat_down₂
 ‘Students did not sit down.’ (Babrazzai 1999, 53 (modified))

¹⁷Strongly (and convincingly) opposed by Dost (2007).

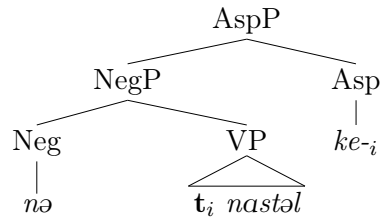


Figure 6.5: Syntactic generation of example (98) (Roberts 2000, 58, modified).

In this case, the “semantically opaque, historical prefix” /ke-/ (Roberts 2000, 56) is moved to ASP. At spell-out, a prosodic inversion to the left subsequently reunites the “proclitic” *ke-* with the following prosodic representation: [ke-[ω [ω nə] nastəl]] (Roberts 2000, 58/59).

While Roberts’ approach provides an almost entirely syntactic solution with prosodic reordering as a last resort, the account is problematic in that all first elements given in any of the Tegey’s word classes *have* to be classifiable as prefixes.¹⁸ While some have been identified as historical prefixes, this is not the case for a majority of the data; and with a certainty, these ‘prefixes’ are not synchronically identifiable.

Another question would be why the perfective aspect is a strong and the imperfective a weak feature. Or why the partial *ke-* is moved completely to Asp in Figure 6.5, although the indication for aspect only comes from stress assignment and not from form (unlike with the perfective marker *wə-* or the perfective/imperfective form of the light verb). And while the classification of the perfective verb construction as two prosodic words (in comparison to the imperfective) points to the right direction, Roberts does not provide a complete phonological account for this specific domain categorization. As will be shown below in Section 6.4.5, the domain of the prosodic word is not sufficient to account for all the data.

Other approaches have focussed on a prosodic analysis. **Hock (1996)** argues at length against a fully syntactic account of 2P clitics, but assumes that “P2 clitics are placed after the first ‘accented element’ (verb, prefix, or first syllable [...]) of an initial verbal clitic group. Elsewhere they are placed after the first (accent-bearing) constituent.” (Hock 1996, 240). Hock does not discuss syntactic placement and the nature of the constituents serving as hosts and does not give a complete account of the prosodic placement, as the focus of his work is on the general discussion if prosodic accounts are applicable for 2P placement at all. However, note that Hock’s statement on the difference between syntactic and prosodic 2P placement in Pashto

¹⁸Or proclitics, as Roberts does not seem to make a clear distinction between the two, although he allows for some proclitics to occupy syntactic projections distinct from the verb (Roberts 2000, 27).

is in line with the fundamental assumption of the proposal presented here.

Anderson (2005) (following mostly Roberts (1997)) considers both, syntactic and prosodic constraints and analyzes Pashto 2P clitics and endoclitics within Optimality Theory. He assumes the constraint ranking as given in (99).

- (99) Integrity(DP), Integrity(PP), Integrity(PPhrase), NonInitial(cl_i , IP)
 >> LeftMost(cl_i , IP)
 (Anderson 2005, 154)

Anderson assumes that every 2P clitic is drawn to the left edge of the IP, but must not be initial, and that the integrity of the major constituents must be preserved. As the verbal complex does not receive an INTEGRITY constraint, the 2P clitic seems to be able to be placed within, but this is not explicitly stated. As to the word-‘internal’ structure, Anderson follows Kaisse (1985) in that he does not concede monomorphemic word status to the a-initial verbs, but assumes that these constructions form two individual prosodic units.¹⁹ Trying to cover all eventualities, Anderson’s analysis is a blend of the syntactic and the prosodic module, also reflected in the OT ranking given in (99) which mixes syntactic and prosodic constraints in that it groups constraints on the integrity of syntactic constituents with constraints on the prosodic placement of 2P clitics.

Interestingly, Anderson also assumes that each stressed element gets assigned prosodic word status and that the phonological phrase constructed on the basis of these prosodic words is the host of the 2P clitics. However, it is unclear, how this translates to the 2P insertion after, e.g., a single prefix, and if that prefix is then considered to be an individual phonological phrase as well.

Dost (2007) proposes an interaction of syntax and prosody to be involved in 2P clitic placement with a strong focus on prosody. His approach is based on word order domains/topological fields in combination with a Head-driven Phrase Structure Grammar (HPSG) architecture (Pollard and Sag 1994). He thus departs from the syntactical analysis and formalizes second position as a linear class, a field, to which only second position clitics may instantiate.

Dost’s main point is based on the fact that 2P clitic placement requires an accented host. Referring to Hock (1996) and Tegey (1977), the conclusion is that unaccented syntactic constituents cannot function as hosts for 2P clitics. As a prosodic host, Dost also identifies the phonological phrase (Dost 2007, 54/84).

Problematic here are cases where the second position clitic is placed after a first accented constituent of a considerable size and how the non-interference of the clitic within such a constituent that has multiple potential hosts can be justified. Furthermore, as with Anderson above, the identification of the host as a phonological phrase

¹⁹The second group of non-discernable verbs in class II is ignored by Anderson.

is also questionable when it comes to verb-partials. This difficulty to determine the correct prosodic domain after which the second position clitic occurs is a common problem with prosodic approaches to second position clitics.²⁰ However, as will be shown below, this tension can be resolved by assuming a mixed approach, where syntactic and prosodic constraints complement each other in the determination of the 2P clitic's final position.

6.3.2 Previous accounts – in LFG

Two previous LFG-related accounts have discussed Pashto endoclitics. **Bögel (2010)** can be seen as an early discussion into the direction of this thesis. One of the main findings was that “while syntax may not intervene in the word-internal structure after the morphological word is formed, prosody still has access to the internal structure of the prosodic word.” (Bögel 2010, 103). However, while the modules were already clearly divided as well, this first approach can rather be seen as a collection of considerations that should be integrated into an analysis, but a concrete LFG analysis had not been fleshed out. Bögel (2010) should thus be seen as an early sketch to the prototype presented in this thesis.

Another recent approach has been proposed by **Lowe (forthcoming)**, who analyses Pashto 2P clitics in the framework proposed by Dalrymple and Mycock (2011) (see Chapter 2, Section 2.6). One important constraint of the Dalrymple/Mycock framework (and a major difference to the approach presented here) is the assumption that the syntactic (s-)string and the phonological (p-)string *have* to be parallel, as they represent two sides of the same string: the s-string is the representation of the ‘surface string’ tokenized into minimal syntactic units, and the p-string is the representation of the ‘surface string’ tokenized into minimal phonological units. While s- and p-string elements are matched up via their cooccurrence in the lexical entries, the individual p- and c-structures are built up independently according to separate principles, thus aiming for strict modular isolation of syntax and prosody. However, if the syntactic (linear) structure is different from the linear order of the elements as determined by prosodic constraints, the parallelity of the p- and the s-string is threatened.

In his paper on Pashto second position clitics (among others), Lowe (forthcoming) develops an approach that allows for the analysis of phenomena like second position clitics and endocclisis in the framework proposed by Dalrymple and Mycock (2011). The underlying assumption is that while the s-string represents linear order and has to be strictly parallel to the p-string, the c-structure represents hierarchical order and is usually based on the s-string. However, if clitics appear in positions that are problematic for the analysis, c-structure is permitted to analyse them in a

²⁰See, e.g., Radonović-Kocić (1996) for similar problems with a prosodic approach to Serbian-Croatian-Bosnian 2P clitics.

more appropriate position. Lowe thus assumes a ‘reordering’ between s-string and c-structure, i.e., during π -mapping. This makes clitic reordering a primarily syntactic affair; repositioning in the prosodic component is impossible.

In his highly constrained formalization of clitic repositioning, Lowe first determines each lexical element as clitic or non-clitic. He develops two functions: $N(\bullet)$ which refers to the ‘next’ element in the linear string, and $N(*)$, which refers to the ‘next’ terminal node in c-structure from a linear perspective. He then states that for any non-clitic word, anything that precedes or follows this element in the c-structure must also precede or follow in the string. (En)clitics are principally not moved, *unless* they appear at the left edge of their syntactic domain. In order to determine this position, Lowe introduces a s-string feature FIELD which includes a set of labels referring to all syntactic categories that dominate the clitic. Lowe then constrains the clitic to the second position in that he states that the preceding s-string element must be a member of the clitic’s FIELD value. These different constraints are then ranked in OT-LFG EVAL to establish the optimal candidate.

In the specific case of Pashto, Lowe (forthcoming, ex.27) proposes the following phrase structure rule to account for syntactic positioning:

(100) $S \rightarrow (XP) (CCL) XP^* (VP)$

So far, neither the phrase structure rule in (100) nor the reordering of clitics in π via the N-function can account for those sentences, where only the 2P clitic and a verb appear.²¹ As Lowe notes, these cases are determined by prosodic constraints. Nevertheless, the repositioning still applies in π , i.e., in the syntax, as the parallelity of p- and s-string has to be preserved under all circumstances. There are several issues with these assumptions in the case of endocclisis.

First, while enclitics still constitute separate syntactic units in the s-string, endocclisis is problematic in that account: As clitic reordering applies within the syntactic constituent and syntactic operations cannot place elements within the morphologically complete word, there is a conflict between the concept of lexical integrity and the empirical data. Furthermore, the endocclitic’s appearance in the s-string (and consequently in the p-string) makes it difficult to match up the string elements with reference to the lexicon. This is especially problematic, if the hosting verb is not separable into distinct morphemes as the individual tokens contributing to the host+endocclitics formation consequently cannot be recognized during the lexical match up.

In order to resolve these difficulties, Lowe argues that the fact that 2P clitics *can* appear between parts of a word is evidence for the syntactic divisibility of that word.

²¹It is not entirely clear why the VP is optional in this account and it is also not clear what would happen if the first optional XP was not given, but if CCL would be followed by another XP before the VP: CCL XP VP. In such a case, the 2P clitic would have to be repositioned after the next major syntactic constituent which might not be identical to the next word/terminal node. Neither $N(\bullet)$ nor $N(*)$ nor the OT EVAL developed by Lowe below can correctly place the 2P clitic. Instead, a third N-function would have to be defined, referring to the sister-constituent of CCL.

Extending Toivonen (2003)'s account of non-projecting categories, Lowe develops an analysis of Pashto verbs in which they contribute not one, but two elements to the s-string and c-structure, where the non-projecting category \hat{V} is syntactically adjoined to V^0 .

$$(101) \quad V \longrightarrow (\hat{V}) \quad V^0 \\ \quad \quad \quad \uparrow=\downarrow \quad \uparrow=\downarrow$$

This (syntactic) complexity is stored as part of the lexical entry. It allows Lowe to project one verb to two terminal nodes and thus to avoid the issue raised above on the placement of a 2P clitic within a syntactically simple element: From a syntactic point of view, the verbal element is now complex.²²

As the placement of the endoclitics is determined by prosodic factors and rearrangement must not take place in p-structure in order to preserve the parallelism of the strings, Lowe relies on OT to determine the correct position of the 2P clitic. There are four constraints that are applied to the candidates in EVAL (Lowe forthcoming):

1. INTERFACE IDENTITY: *no rearrangement of material is permitted at the syntax-phonology interface.* This is an inviolable constraint that applies to the p-string/s-string interface and prohibits a change in linear order.
2. *MOVE-LFT: *Do not move a clitic leftwards in the π mapping.* This constraint applies to the rearrangement of clitics between s-string and c-structure.
3. ALIGN(VP,R;S,R): *Align the right edge of the VP with the right edge of the clause.* This purely syntactic constraint enforces the verb-final position.
4. ALIGN(CL,L;PW,R): *Align the left edge of a prosodic clitic with the right edge of a prosodic domain.* Here, the placement after the first stressed element (the PW) is ensured.²³

These four OT constraints are then ranked by Lowe as shown in (102):

$$(102) \quad \text{INTERFACE IDENTITY, ALIGN(VP,R;S,R), ALIGN(CL,L;PW,R)} \\ \quad \gg \text{*MOVE-LFT} \\ \quad \quad \quad \text{(Whereby (at least) INTERFACE IDENTITY is inviolable and should thus not be ranked} \\ \quad \quad \quad \text{together with the alignment constraints, but should be on the top of the scale.)}$$

²²In principle, Lowe thus assumes 'mesocclisis' (See Chapter 5, Section 5.1.1) instead of 'endocclisis' for Pashto.

²³Note that this constraint is problematic in several respects: First, 2P clitics are not the only clitics in Pashto, but there is a second group of preverbal proclitics, which would be placed incorrectly. Second, the notion CL is used in both, the syntactic and the prosodic description of the lexical item, but strictly speaking, 'clitic' is not something that phonological processes can understand. Rather, it should be 'a prosodically deficient item that looks for a host to its left'. However, both issues can be resolved with a more fine-grained description

Note that these constraints apply to different ‘positions’ in the overall architecture, i.e., an interface constraint is paralleled with a syntactic (c-structure) and a prosodic constraint all of which precede a syntactic constraint that applies during the π -projection. A clear separation of modules is thus no longer given in OT EVAL. Example (103, repeated from (90)) illustrates the complete process.

- (103) $\text{ʔák} = \mathbf{me} \text{ w}\alpha\text{h}\text{ə}$
 shake₁ I shake₂
 ‘I shook it.’

The lexical entry for the verb is given in Table 6.4.

$\text{ʔák w}\alpha\text{h}\text{ə}: \check{V} V$ (↑ PRED)=‘shake’ (↑ ASPECT)= ‘perf’ ----- /ʔak _s / /wα _w .hə _w /	$\text{ʔák w}\alpha\text{h}\acute{\text{ə}}: \check{V} V$ (↑ PRED)=‘shake’ (↑ ASPECT)= ‘imperf’ ----- /ʔak _w / /wα _w .hə _w /
---	---

Table 6.4: Lexical entries as proposed by Lowe (forthcoming, 37, ex. 35) for *ʔakwαhə*.

In Table 6.4, the top part corresponds to the syntactic form and the bottom part to the phonological form. Based on the assumption that each prosodic word must contain one stressed syllable, the first ‘part’ of the p-form entry for the perfective aspect is given prosodic word status as it carries main stress. In the imperfective, on the other hand, the main stress is on the last syllable. Thus, the complete structure is united under one prosodic word. Figure 6.6 shows the complete architecture as proposed in Lowe (forthcoming) by means of the perfective example in (103), which is the optimal candidate selected by the constraint ranking given in (102).²⁴

²⁴The complete OT analysis is left out for space reasons. The interested reader is referred to Lowe (forthcoming, 39, ex40).

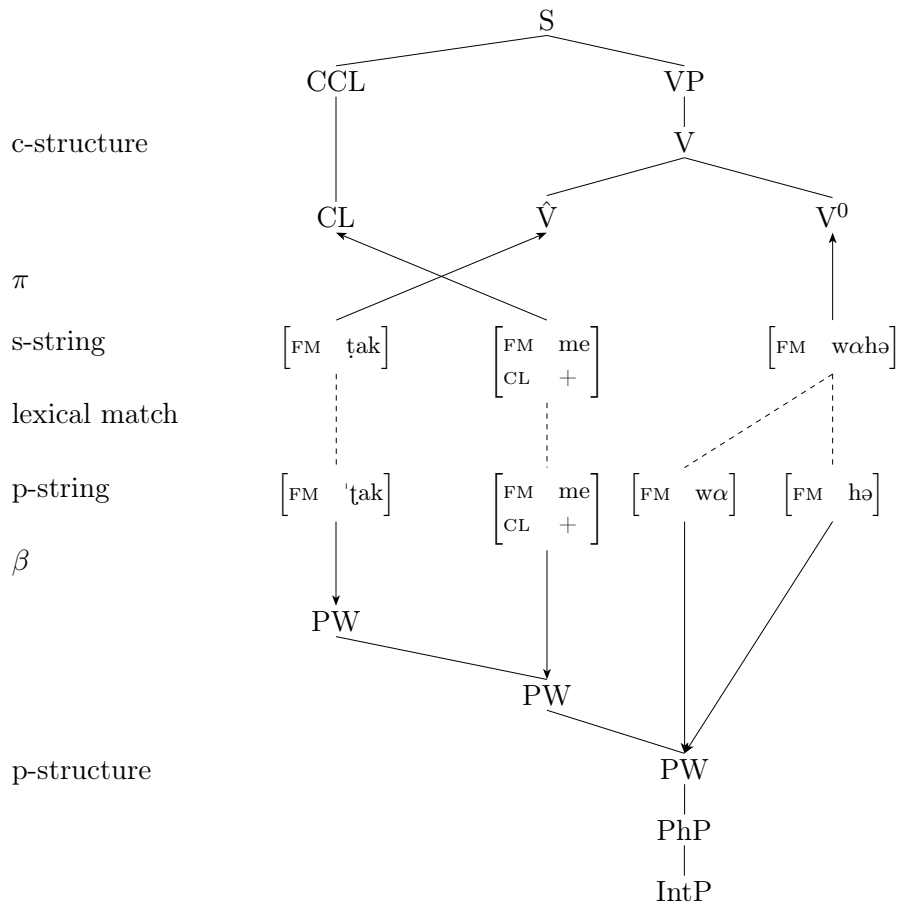


Figure 6.6: Analysis of (103) as proposed by Lowe (forthcoming, 39, (40) (modified)).

In Figure 6.6, the 2P clitic *me* cannot occupy its originally initial c-structure position in the surface string, as it needs to lean on a preceding host. Consequently, it is placed after the first prosodic word in the related (s-)string.

While a rearrangement of 2P clitics in the π -projection is a possible solution to deal with syntactically determined 2P clitic positioning,²⁵ Pashto 2P clitics are partly determined by prosody. However, the inviolable constraint that p- and s-string are strictly parallel forces Lowe (forthcoming) to analyse Pashto 2P clitics mainly within the syntax. Thus, this account is problematic with respect to modularity, as a rearrangement of 2P clitics in the syntactic constituent *on the basis of prosodic constraints* is difficult to justify. The resulting problems, e.g., the interruption of syntactic atoms, is evaded by the assumption that Pashto verbs are syntactically complex. However, the fact that 2P clitics can intrude into the verb is *not* a justification for this assumption, especially not if prosody is the triggering factor (and not

²⁵See Lowe (forthcoming) for an account of 2P clitics in Ancient Greek.

syntactic features).

As Lowe (forthcoming) stresses repeatedly, the approach is modular in the sense that each module “exclusively and exhaustively analyses one aspect of linguistic structure, and that no aspect of linguistic structure is analysed, even marginally, in any module except the one specific to it.” (Lowe forthcoming, 14). This strict separation allows for no communication between the modules which forces Lowe to resort to grammar-external mechanisms to allow for an interaction between the modules: Optimality Theory. As Lowe states, “[...] the EVAL assesses candidates in relation to constraints on all the different components of the grammar, not only syntax but also prosody. In a sense, then, OT EVAL functions as the locus for interaction between syntax and prosody, which in the modular grammatical architecture itself are kept strictly separated.” (Lowe forthcoming, 33). That is, after the strong focus on strict modularity has not yielded the correct results, additional formal power is applied, where the optimal candidate is determined by a ranking of constraints mixed together from both modules, which could be viewed as a violation of modularity through the backdoor.

In the following sections, an alternative proposal is given without having to revert to external mechanisms to account for modular communication. By assuming that the syntactic string is in most cases parallel to the linear order of the phonological string, but that prosodic constraints can lead to a mismatch of both, Pashto 2P clitics in the verbal complex can be explained in purely prosodic terms, while the preverbal placement can be analysed in purely syntactic terms.

6.4 2P clitics in Pashto: a new proposal

Pashto second position clitics can occupy a seemingly conflicting variety of positions in the string: On the one hand, 2P clitics seem to follow the first syntactic constituent which they cannot interrupt. On the other hand, 2P clitics also seem to be sensitive to stress. If the ‘first syntactic constituent’ is destressed, the clitic is placed after the next constituent that carries stress. And finally, if the arguments are dropped and the verb is clause-initial, 2P clitics are placed depending on an aspect-caused stress shift: *after* the verb in the imperfective and *within* the verb in the perfective aspect.

The resulting conflict is twofold: a) Why is the 2P clitic placed after the first syntactic constituent (no matter the size and stress distribution) in one case, but can be placed even within the verb in other constructions? And b), as the placement is also determined by stress, what is the related prosodic domain for 2P clitic placement if a variety of hosts ranging from a major size down to partial words have to be considered? As the literature review of the previous section shows, these questions have not been answered satisfactorily with respect to the complete data set.

This section proposes a new solution based on several syntactic and phonological considerations. Second position clitics in Pashto are placed according to two clearly separated constraints: prosodic and syntactic. Pashto second position clitics are

first and foremost syntactic clitics that are strictly placed after the first (preverbal) constituent of the clause. However, consider the following example (repeated from (87)), where the 2P clitics are not placed after the destressed preverbal postpositional phrases, but after the verb, i.e., the first element carrying stress.

- (104) rα= ta= pe= gαndá =de
 me for by_him sew you
 ‘You were having him sew it for me.’ (Tegey 1977, 118)

The resulting conclusion was that the first syntactic constituent, after which the 2P clitic attaches, *has* to be stressed. Consequently, every modular account had to find a prosodic domain that could include major syntactic constituents as well as single syllables which proved to be rather impossible. However, as the following sections will show, this search for a common prosodic domain is irrelevant, because

There are no unstressed constituents before the verbal complex in Pashto!

In other words, any unstressed material *is part* of the verbal complex (henceforth VC), which is always final. More concretely, in contrast to their fully stressed counterparts, the preverbal clitics in (104) are not syntactic constituents that are immediate daughters of S. Instead, they (syntactically) become part of VC and are thus beyond the range of the syntactic positioning of 2P clitics, which are always syntactically placed preceding the verbal complex. Syntactic and phonological evidence for this claim will be given below.

2P clitics are thus never left with an unstressed preceding constituent. Once the 2P clitics are stranded in the preverbal position without an adequate host, prosodic inversion is applied as a last resort, i.e., the 2P clitics are placed *according to prosodic constraints*. The prosodic host is the first prosodic word in a nested prosodic word structure that stretches over the VC and its members, where the right boundary of the ‘first prosodic word’ is determined by the stressed element in the verbal complex. Evidence for this proposal comes from several strictly ordered postlexical phonological processes that are bound by the domain of the (nested) prosodic word: vowel coalescence, vowel harmony and initial /k/-deletion in complex verb structures.

The following sections will first focus on the data that determines syntactic second positioning and an implementation thereof in LFG. After the syntactic frame has been established, the discussion will focus on the prosodic placement of second position clitics within the verbal complex.

6.4.1 The syntactic structure of Pashto

Many researchers describe Pashto as essentially an SOV language which allows for scrambling under the right conditions. Scrambling is thus possible “as long as the grammatical functions of the arguments are clear from context or case-marking”

(Roberts 2000, 13), but it has also been noted that scrambling can involve slight differences in meaning (Tegey 1979, 379). Furthermore, scrambling does not involve the verb and its associated elements: The verb, part of the verbal complex VC, is always final in Pashto (Tegey and Robson 1996, 166).

Pashto also allows for subject/object drop if the dropped NP is non-focused, non-contrastive and non-conjoined (Tegey 1977, 12). As a consequence, constructions where the verbal complex is the only element left are possible as well. These reduced constructions and their interaction with 2P clitics will be discussed in more detail in Section 6.4.4. As the pre-verbal-complex constituents in the Pashto clause can scramble freely, it is assumed that Pashto has a flat syntactic structure in that all major constituents are immediate daughters of S (cf., Mohanan 1990, Butt 1995, for Hindi/Urdu).

2P clitics are placed after the first major constituent of a phrase, which can be the subject, the object, an adverbial phrase, a postpositional phrase, etc. Furthermore, there is no limit to size: The constituent can be of a considerable size in one case ((105a)), and consist of one word (or even a partial word, see Section 6.4.4) in another case ((105b)).

- (105) a. [aḡa šəl kaləna danga aw xəysta peḡla] =**me** nən byα wəlida
 that 20- year tall and pretty girl I today again saw
 ‘I saw that 20-year old tall and pretty girl again today.’ (Tegey 1977, 83)
- b. [nən] =**me** mele ta byayi
 today me picnic to takes
 ‘Today he takes me to a picnic.’ (Tegey 1977, 84)

Another criterion for the 2P clitic’s host which has been repeatedly stated in relation to the ‘first major constituent’ is that it has to be stressed. As Tegey (1977, 112) notes, the constituent has to carry some kind of lexical stress in order to function as a host: The 2P clitics are thus “placed after the first major surface constituent that bears at least one main stress – where ‘major constituent’ may be directly dominated by S, VP, or V” (Tegey 1977, 122). As briefly stated in the previous section, these restrictions have led to some confusion as it requires an explanation as to why the major syntactic constituent in, e.g., example (105a) cannot be interrupted on the one hand, while the 2P clitic can be placed within the verb on the other hand. The attempt to find a common prosodic denominator has been relatively futile.

The unstressed ‘constituents’ that have been taken as a basis for the claim of a ‘stressed major constituent’ are the preverbal clitics as exemplified (104). This thesis claims that there are no unstressed constituents in Pashto that are immediate daughters of S, but that the preverbal clitics are part of the verbal complex. As a consequence, syntactic and prosodic placement of second position clitics can be neatly separated and the analysis becomes straightforward, in that second position clitics in Pashto are syntactically placed, but are subjected to prosodic inversion

once stranded in clause-initial (i.e., the intonational phrase-initial) position. In the following section, this claim will be supported by syntactic, and later by phonological data.

6.4.2 The syntax of preverbal clitics

In addition to the 2P clitics introduced above, there is also a group of proclitics which consist of (optional) weak oblique pronouns in combination with postpositions.²⁶ These items originally correspond to full post-positional phrases and are immediate daughters of S. However, if reduced to their cliticized status, the elements are placed directly in front of the verb, with the verb carrying main stress. In (106), this process is illustrated, where the strong pronoun *mα* ‘me’ in (106a) is found in its weak counterpart *rα* and is placed directly preceding the verb ((106b)). Any other position would render the sentence ungrammatical.

- (106) a. construction with a **strong** oblique pronoun: *mα*
 tor [*mα sara*] ɖer ɣə pezani
 Tor me with very well acquainted
 ‘Tor is very well acquainted with me.’
- b. construction with a **weak** oblique pronoun: *rα*
 tor ɖer ɣə [*rα= sara=*] pezani
 Tor very well me with acquainted
 ‘Tor is very well acquainted with me.’ (Tegey 1977, 228)

It is not clear why the preverbal clitic in (106b) is no longer present in the original syntactic position as in (106a). Its placement next to the verb in (106b) cannot be determined by purely prosodic terms, as the preverbal clitic in (106a) would find host-‘material’ on each side. One hypothesis would be that these clitics are ‘syntactic clitics’,²⁷ in that they do not only attach prosodically to their host, but also syntactically. In Pashto, the syntactic placement of the preverbal clitics in the verbal complex would ensure continuing functional scope over the complete clause, while a syntactic attachment to the ‘next constituent’ could possibly strand the clitics’ functional information within the functional scope of that constituent (e.g., the subject, the object, an adjunct). However, the exact reasons for the preverbal clitic placement have to be left for further research.

This difference in stress between the postpositional phrases in (106a) and (106b) is also reflected in the following examples. In (107a), the postpositional phrase *layla na* is lexically stressed and can thus function as a host for the following 2P clitic *de*. In (107b), on the other hand, the postpositional phrase is a contracted and

²⁶As this group is in itself not relevant to the discussion of 2P clitics except for their position within the verbal complex, the description here suffices. For more information on this group of proclitics, see Tegey (1977, 214ff.).

²⁷Anderson (2005)’s Type 2 clitics, see Chapter 4, Section 4.1.1.

unstressed preverbal clitic. The 2P clitic *de* no longer finds a suitable host in *tre* and thus encliticizes to the following verb.

- (107) a. [laylɑ na]_{PP} =**de** [axistə]_{VC}
 Layla from you buy
 ‘You were buying it from Layla.’ (Tegey 1977, 114)
- b. [[tre=]_{PP} axistə]_{VC} =**de**
 from_her buy you
 ‘You were buying it from her.’ (Tegey 1977, 114)

Supported by (106), it is assumed that the examples given in (107) are fundamentally different in their syntactic representation. While in (107a), the postpositional phrase inhabits a separate daughter node of S, the reduced form *tre* in (107b) is a preverbal clitic and is part of the verbal complex. This results in two different syntactic c-structure representations:

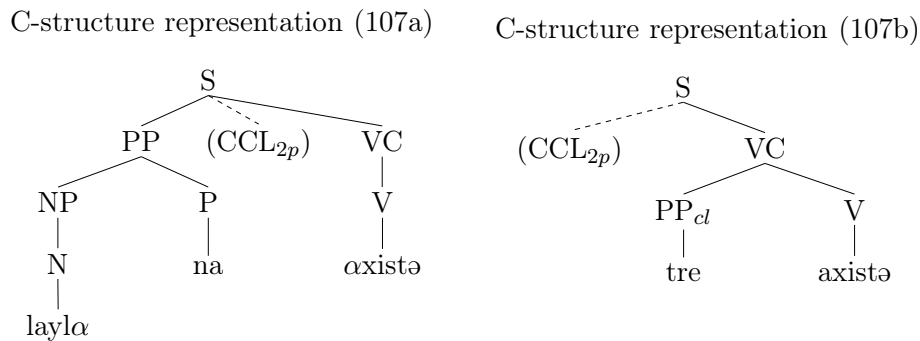


Figure 6.7: C-structure representations for (107a) and (107b).

While the postpositional phrase in (107a) provides the environment for syntactic 2P placement, the preverbal clitics in (107b) have been placed in the verbal complex. 2P clitics are stranded in the position preceding the verbal complex without an adequate host (Figure 6.7b). Consequently, they will be subject to prosodic inversion within p-structure.

Note that any destressed constituent (i.e., which becomes a clitic), will either become part of the 2P clitic cluster (e.g., weak pronouns) or it will be part of the verbal complex. In other words, there will *never* be an unstressed constituent that is an immediate daughter of S. The claim that the major syntactic constituent, to which 2P clitics attach if they precede the verbal complex, *has* to be stressed is thus invalid, because the contrary situation never occurs. Consequently, the discussion about a common prosodic domain for giant-sized constituents and parts of the verb is irrelevant as well, because the 2P clitics are placed *according to syntactic constraints* if preceding the verbal complex. Once the 2P clitic is left in the syntactically initial

clause position, it will be subject to postlexical prosodic inversion, as it is also placed initially in the intonational phrase, but requires a preceding host. It is only then that a determination of a prosodic domain is relevant, as it is a necessary requirement for the postlexical replacement via prosodic inversion. Thus, syntactic and prosodic 2P placement can be neatly separated in the approach presented here. As the following sections show, the analysis is straightforward.

6.4.3 An analysis of syntactic 2P clitics in LFG

From the data presented above it can be established that second position clitics occur after the first major syntactic constituent, unless the clause is reduced to the verbal complex. In this case, as will be discussed in Section 6.4.4, second position clitics are subject to prosodic inversion.

Within *c*-structure, possible 2P clitics sequences are incorporated under one common node: CCL, which was coined by Bögel et al. (2010) to express 2P clusters in Serbian-Croatian-Bosnian. The CCL node is an immediate daughter of S and has no functional annotation except for $\uparrow=\downarrow$, thus allowing for any information provided by the individual clitics to reach the sentential functional scope. The advantage to a more fine-grained representation where each 2P clitic is individually attached to the S node is the treatment (and placement) of templatic 2P clitic sequences as a single unit. Therefore, in the following definitions, CCL will be used as a representation of all possible 2P clitic sequences as they have been described in Section 6.2.

Figure 6.8 shows an abstract *c*-structure representation of the Pashto clause, where CCL and VC are immediate daughters of S ($M^{-1}(S)$).

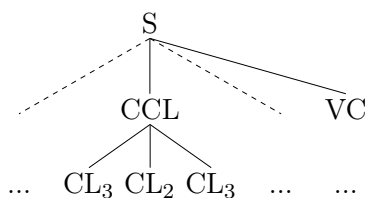


Figure 6.8: Abstract representation of the Pashto clause.

There are two options for the 2P clitic to appear in the *syntactic* linear order: After the first major constituent or, if no major constituent is present, clause-initially preceding the verbal complex. The corresponding *c*-structure rule encoding both options for the occurrence of 2P clitics is shown in 108.²⁸

²⁸The overall *c*-structure rule for S will of course be much larger. The example shown here is reduced to the two options ([1 | 2]) that occur with 2P clitics (CCL).

$$(108) S \longrightarrow \dots$$

$$\begin{array}{l} | \{XP \text{ CCL } XP^*\} \\ | \text{CCL}\} \\ \text{VC } | \end{array}$$

where $XP = \{NP \mid PP \mid AP \mid \text{AdjP}\}$

Note that the c-structure rule in (108) provides two possible placements for CCL. The first part describes the placement in the second syntactic position after the first major constituent. The other option, where CCL is the only other element left next to the verbal complex, is given in the second part of the rule. Both options are followed by the verbal complex VC, thus resulting in two possible linear structures:

1. XP CCL XP* VC
(Option 1, where the clitic is placed in second position after the first major constituent. No prosodic inversion is required, as the first major constituent always carries lexical stress.)
2. CCL VC
(Option 2, where the clitic is the only element left next to the verbal complex. This option provides the environment for prosodic inversion, as the clitic will be in the initial position within the intonational phrase in p-structure.)

Note that this approach adheres to strict modularity, as the syntactic placement does not require any reference to the phonological properties of its elements. Instead, the prosodic placement is determined in p-structure on the basis of the syntactic output, which is unproblematic with Option 1, but which crucially will place the 2P clitics in a prosodically unacceptable position in Option 2.²⁹

²⁹Note that there is an alternative to the phrase structure rule given in (108). Following Bögel et al. (2010), the second syntactic position can also be formalized as follows:

$$(109) \textit{Second}(x, y) \equiv x/y \cap [[\Sigma - y] y [\Sigma - y]^*] \quad (\text{Bögel et al. 2010, Fn15})$$

$\textit{Second}(x, y)$ is defined as the insertion of y in the second position of every string in a regular language x . The righthand side thus can be read as: Insert y into the regular language x (x/y), where y is constrained (intersected with, \cap) to a position after an element that is not y ($\Sigma - y$) and possibly followed by 0 or more elements ($*$) that are not y .

In the specific case of Pashto, x is equivalent to RHS_s which “denotes the possible expansions of the clausal S node” (Bögel et al. 2010, 117), y is equivalent to CCL, and any daughter node of S ($M^{-1}(S)$) is equal to Σ . The placement of syntactic second position clitics in Pashto can thus be defined as follows:

$$(110) \textit{Second}(\text{RHS}_s - \text{VC}, \text{CCL}) \equiv \text{RHS}_s / \text{CCL} \cap [M^{-1}(S) - \text{CCL}] \text{CCL} [M^{-1}(S) - \text{CCL}]^*$$

While the disjunction given in (108) needs less machinery and is relatively easy to implement into a computational grammar, the formalization given in (110) reflects the language from a theoretical perspective: Pashto allows for scrambling before the verbal complex; the daughters of S thus do not have a fixed position for functional reasons (unlike, e.g., English, where the subject *John* in sentences like *John kissed Mary* can appear nowhere else). However, the CCL node is different in

Summing up, Section 6.4.1 gave a brief introduction to the syntactic structure of Pashto. Several cases where 2P clitics were placed in the second position after the first major constituent were discussed. The findings were transformed into a phrase structure rule ((108)) that allowed for the placement of 2P clitics after the first major constituent (i.e., after the first immediate daughter) of S. While the syntactic placement of the 2P clitics can be completely described in terms of the phrase structure rule given in (108), the prosodic placement of the 2P clitics within the verbal complex was only briefly touched upon by the second part of the rule: $S \rightarrow \text{CCL VC}$. These cases, where the 2P clitic is left with only the verbal complex in a clausal structure will be discussed in the following, where it will be shown that, in contrast to the purely syntactic account presented above, the 2P clitics are placed on purely prosodic grounds within the verbal complex.

6.4.4 Prosodic second position: the verbal complex

The following section focuses on the (syntactic) constructions where the 2P clitics are left with nothing but the verbal complex to attach to: $[\text{CCL VC}]_S$. In contrast to the syntactic account of second position clitics in sentences with preverbal syntactic constituents discussed in the previous section, the underlying assumption made here is that the 2P clitics are subject to postlexical prosodic inversion once they are stranded in the position preceding the verbal complex. Furthermore, it will be shown that the ‘second position’ within the verbal complex is no longer defined by syntactic means, but that the positioning of the 2P clitics is solely determined by prosodic constraints.

One important question that should be discussed in this context is the prosodic domain that the 2P en(do)clitics attach to. In order to determine the answer to this question, several phonological processes will be described and discussed in the following sections: vowel coalescence, vowel harmony, and deletion of initial /k/ in complex predicates.³⁰ Each phonological process adds a piece to the puzzle of prosodic placement and consequently provides a deeper insight into the prosodic phrasing within the verbal complex on the one hand and into the linear succession of these postlexical phonological processes on the other hand.

However, before these issues can be described any further, a brief description of the verbal complex and its members is in order, as the interaction of the different members among each other and in relation to the phonological processes adds to

this respect. It *must* be placed in syntactic second position, but for no obvious *functional* reason. Its placement thus seems to be determined by different constraints compared to the ‘standard’ XPs. This difference is, in some way, expressed by *second(x,y)*, but not by the disjunction in (108).

³⁰Note that there are several other phonological processes applying within the verbal complex, e.g., labialization and contraction. However, as these phenomena do not add to the discussion of prosodic phrasing or duplicate the findings provided by the processes discussed here, they were left out for reasons of space and simplification. The reader is referred to Tegey (1977) for more information.

the overall solution of the puzzle. A complete analysis of all possible processes and combinations goes beyond the scope of this chapter. The following section was thus reduced to the optional members of the verbal complex that are essential for the following analysis of endoclitisis from a phonological perspective: In addition to the verb itself and the group of 2P clitics discussed above, this includes the preverbal proclitics and the negative particles.

6.4.4.1 Preverbal proclitics

The preverbal clitics have been briefly introduced in Section 6.4.2 as a group of proclitics that consist of a combination of weak oblique pronouns and postpositions. It has been argued above that while the full postpositional phrase is an immediate daughter of S, the reduced/destressed version becomes a member of the verbal complex.

If the preverbal clitics are the only items left next to the verb, they will never be placed after the first stressed element, but are strictly placed in front of the verb (unlike 2P *enclitics*).³¹ Thus, in (111), the preverbal clitic is not placed after the stressed perfective prefix *wə-*, but in the preceding position. The preverbal clitics can thus be placed clause-initially: they are proclitics whose position will not change whether they appear in an initial position or in a medial clause position preceding the verb.

- (111) *pe= wədaredə*
 on_it PERF.stood
 ‘He stood on it.’ (Tegey 1977, 240)

As has been shown in (107b) above, the preverbal clitics cannot function as hosts for the 2P clitics. They are ‘ignored’ and the 2P clitics are placed after the first stressed element (the perfective prefix in the (112)).

- (112) *rə= ta= pe= wə =de gəndə*
 me for by_him PERF you sew
 ‘You had him sew it for me.’ (Tegey 1977, 118)

It can thus be concluded that the preverbal clitics are placed before the verb and its prefixes and can also inhabit the clause-initial position, but they cannot serve as a host for 2P clitics themselves. The following section introduces another member of the verbal complex, the negative marker, and shows (among other findings) how this marker interacts with the preverbal clitics.

³¹For rare exceptions caused by constraints at the syntax-information structure interface, see Tegey (1977, 239ff). In these cases, the ‘preverbal’ clitics can be placed postverbally.

6.4.4.2 Negative marker

Pashto has two negative particles: *nə*, which is used in ordinary negation, and *mə*, used for negative commands. Both particles have identical behavior, which is why they will not be treated separately in the following section. As it is the case with the perfective prefix, the negative particles always carry the main stress in constructions involving only the negative particle and the verb ((113b), in comparison to (113a)).

- (113) a. gaḍīgi
 ‘She is dancing.’
 b. nə gaḍigi
 ‘She is not dancing.’ (Tegey 1977, 111)

The assignment of stress predicts the negative marker to qualify as a host to the 2P clitics and indeed, if second position clitics occur in a sentence that has been reduced to the verbal complex, they will be placed *after* the stressed negative marker ((114a)), but only if no other acceptable host is present outside of the verbal complex ((114b)).

- (114) a. [nə =ba gaḍigi]VC
 not should dance
 ‘You should not dance.’ (Tegey 1977, 111)
 b. tə =ye [nə ešawe]VC
 you it not boil
 ‘You aren’t boiling it.’ (Tegey 1977, 149)

The negative marker is also subject to vowel coalescence as it was described in Section 6.2.3.4: The final /ə/ of the marker coalesces with verb-initial /a/, forming [a]. If a 2P clitic intervenes, vowel coalescence still takes place ((115b)).

- (115) a. tə =ye [nəxla]VC *tə ye nə axla
 you it not buy
 ‘You are not buying it.’ (Tegey 1977, 149)
 b. [nə =ye xle]VC *nə ye axle
 not it take
 ‘You are not taking it.’ (Tegey 1977, 176)

The negative particle thus shares several common characteristics with the perfective marker and the question arises if the negative particle should be treated as a prefix as well. However, as the following examples show, the negative marker also exhibits several properties that mark it as an individual syntactic item. In contrast to the perfective marker, the negative marker *precedes* preverbal clitics ((116) *contra* (112)).

- (116) tor nə pe= wahi
 Tor not with it hits
 ‘He is not hitting Tor with it.’ (Tegey 1977, 236)

As the preverbal clitics are not prosodically deficient enclitics that need to encliticise to a host on their left, but are generally placed as closely to the verb as possible, there is no indication for prosodic replacement of the preverbal clitics after the negative marker. Thus, the construction in (116) can be taken as an indication that the negative marker is syntactically placed externally to the perfective marker and the preverbal clitics and that it is not part of the morphological generation of the verb form, but a syntactic element in its own right.

Another aspect of the negative marker is that it is always placed before the verb in the imperfective ((117a)). However, it behaves similarly to second position clitics in the perfective aspect in that the marker is placed after the first stressed element ((117b and c)).

- (117) a. šagerdan [nə kenastəl]_{VC}
 students not sat_down
 ‘The students were not sitting down.’ (Babrakzai 1999, 53 (modified))
- b. šagerdan [ke nə nastəl]_{VC}
 students sat_down₁ not sat_down₂
 ‘Students did not sit down.’ (Babrakzai 1999, 53 (modified))
- c. asad tor [wə nə pezəni]_{VC}
 Asad Tor PERF not knew
 ‘Asad didn’t know Tor.’ (Tegey 1977, 171)

The reason for this behavior, which Tegey called “Negative Reordering” (Tegey 1977, 172), is unclear. A full analysis of the challenging data on the negative marker is beyond the scope of this thesis.³² The important fact to note for the current

³²One hypothesis could be that the formerly stressed negative marker is destressed in the immediate presence of another stressed element (the perfective marker/the first partial of the perfective verb) in order to avoid a stress clash (similar to the English rhythm rule (Gussenhoven 1991, Shattuck-Hufnagel et al. 1994)). As a consequence, the negative marker becomes prosodically deficient and is thus subject to prosodic inversion to place it in a position that allows for it to lean on a host. However, while Tegey (1977, 175/177) describes the inverted negative marker as ‘unstressed’, Tegey and Robson (1996, 26) specifically mention that *wə* and *nə* are stressed next to each other. My own native speaker recordings show this to be true, and as a consequence, the destressing hypothesis can probably be discarded.

A second hypothesis is that the negation marker inhabits a specific position within the prosodic domain of the verbal complex. This position is the first one in the verbal complex if the negative marker carries the nuclear stress. However, if another element within the overall domain carries main stress as well, the negation marker will inhabit the second position after the first element carrying stress. So far, as will become clearer below, this hypothesis has not been falsified, but the topic awaits further research.

discussion is that the negative marker, although it is stressed, can be subject to ‘reordering’, a process that is usually associated with 2P clitics.

A third interesting characteristic of the negative marker is its behavior with respect to the phonological process of vowel coalescence. As noted above in example (115), negation seems to follow the behavior of the perfective marker with respect to vowel coalescence. However, as (118) shows, the negative marker is *not* subject to vowel coalescence if a preverbal clitic intervenes.³³

- (118) $nə$ te= axle * $n\alpha$ te xle
 not from_him take
 ‘... are not taking ... from him (Tegey 1977, 176)

Furthermore, the negative marker is also not subject to vowel coalescence in a case of negative reordering ((119)). Note also that negative reordering only applies within the verbal complex — the negative particle does *not* copy the syntactic placement of 2P clitics and negative reordering is also *not* restricted to constructions where the verbal complex is the only constituent in a clause.

- (119) $tə$ =**ba** as [$w\alpha$ $nə$ xle]_{VC} * $tə$ ba as $wə$ $n\alpha$ xle
 you will horse PERF not buy
 ‘You will not buy the horse.’ (Tegey 1977, 172)

In (119), The perfective marker undergoes vowel coalescence with the initial vowel /a/ of the verb: $wə+axle \rightarrow w\alpha xle$. After vowel coalescence has applied, the negative marker has been placed between the prefix and the verb stem; any other combination is ungrammatical.

A last indicator for the claim that the negative marker is syntactically independent while the perfective marker is a prefix of the verb comes from the Pashto script: While the perfective marker is written together with the verb, the negative marker is an orthographic token by itself.

6.4.4.3 Interim summary

On the grounds of the examples presented above, so far, several crucial facts can be established about the members and the phonological processes of the verbal complex, which are important for the following analysis of the host for 2P clitics.

³³As vowel coalescence is a very early postlexical process which applies even before 2P clitics are repositioned, (117) could also be taken as further evidence that the preverbal clitics are not positioned by postlexical prosodic constraints, but are placed within the verbal complex syntactically. Otherwise, postlexical prosodic placement of preverbal clitics would have to take place before vowel coalescence and 2P placement.

1. *Several linear orders are valid for the verbal complex.*³⁴

neg	2P	preverbal clitics	perf partial ₁	2P	neg	verb	2P
		(✓)	(✓)			✓	
✓	(✓)	(✓)				✓	
		(✓)	✓	(✓)	(✓)	✓	
		(✓)				✓	(✓)

Table 6.5: A selection of possible linear orders of elements within the verbal complex.

2. *The negative marker is an independent syntactic item.*

This conclusion is supported by several facts: First, the negative marker precedes the preverbal clitics ((118)), where the perfective marker follows them ((111)). Second, the negation marker is subject to “negative reordering” if another element within the verbal complex carries main stress. In these cases, the negative marker seems to be subjected to the same ‘final destination’ as the 2P clitics within the verbal complex, albeit it not being prosodically deficient and the presence or absence of major constituents outside of the verbal complex being irrelevant. Third, the perfective marker has precedence over the negative marker in cases of vowel coalescence ((119)). And last, the negative marker is written independently in the Pashto script (as opposed to the perfective marker).

3. *Vowel coalescence is a postlexical phonological process.*

This conclusion derives from the fact that the negative marker can be subject to vowel coalescence ((115)). If vowel coalescence only occurred with the perfective prefix, the question would remain if it was simply a *lexical* phonological process.

The following section discusses the possible prosodic phrasing of the verbal complex. In addition to vowel coalescence, other postlexical phonological processes will be introduced that provide further insights into the overall aim to determine the prosodic host for the 2P clitics.

6.4.5 Phonological processes within the verbal complex

The previous sections have discussed the phonological process of vowel coalescence with respect to the perfective prefix and the negative marker. It was established that the placement of 2P clitics takes place *after* the vowel coalescence has applied

³⁴This does not include ALL valid linear orders, but the ones that are relevant for the overall aim which is to find a pattern for the placement of second position clitics within the verbal complex. One factor is the ‘host’ to which the 2P clitics attach. In order to be able to apply the upcoming phonological rules that are necessary for the identification of this host, the linear order of these elements and their interaction is crucial.

((115)).³⁵ Furthermore, it was established that vowel coalescence is a postlexical process that can apply to syntactically independent negative markers as well as to the perfective prefix; thus, vowel coalescence can ‘cross’ prosodic constituent borders (presumably below the prosodic word). The remaining question is if vowel coalescence is restricted to a particular prosodic domain. As the following example shows, this is, indeed, the case.

- (120) kor špənə axli *špənə xli
 house shepherd buys
 ‘The shepherds are buying the house.’ (Tegey 1977, 148)

In (120), vowel coalescence cannot apply between two lexical words where each bears an individual lexical stress. It is, however, difficult to identify the prosodic border between the verb and ‘shepherds’ at this point, as it could be the prosodic word boundary or, e.g., a phonological phrase boundary. It is just as difficult to establish the overall prosodic domain of the verbal complex on the grounds of vowel coalescence alone. To gain further insight, the next section discusses another phonological process: regressive vowel harmony.

6.4.5.1 Vowel harmony

Regressive vowel harmony in Pashto is a postlexical phonological process whereby the presence of the high vowels /i/ and /u/ raises the mid-vowels /o/ and /e/ in a preceding syllable to high: *tok-i* → *tuk-i* (‘joke’). Vowel harmony applies to other word classes as well, but cannot spread across word boundaries ((121)).

- (121) ɣe wuxə *xi wuxə
 good camels
 ‘Good female camels’ (Tegey 1977, 159)

In terms of prosodic phrasing of the verbal complex, these findings are especially interesting. In (121), the two words form a syntactic unit in that the adjective modifies the noun. In terms of prosodic structure, (121) would be interpreted as a phonological phrase. As vowel harmony cannot spread between these two fully stressed words, it can be concluded that vowel harmony is restricted to the domain of the prosodic word, and not to the phonological phrase. If, on the other hand, vowel harmony would be restricted by the phonological phrase, then it could be expected that constructions like (121) would display vowel harmony as well.

In contrast to (121), vowel harmony applies between host and 2P clitic within the verbal complex, as the following two examples show.

³⁵The same is true for negative reordering, but a full analysis of this phenomenon is beyond the scope of this thesis.

- (122) a. wə =**di** (*de) guri
 PERF should see
 ‘He should see him.’
- b. nə =**mi** =**di** wini
 not me should see
 ‘He shouldn’t see me.’ (Tegey 1977, 168)

As Tegey notes, vowel harmony in 2P clitics “arises only when they are placed after an item that itself makes an accentual unit with the following verb root [...] such that the total construction [...] constitutes a single accentual unit.” (Tegey 1977, 167/168). And indeed, if the 2P clitic is placed following a previous syntactic constituent outside of the verbal complex, vowel harmony will not spread from the verb to the 2P clitic, even if the two are adjacent.

- (123) patang =**me** (*mi) [wini]_{VC}
 Patang me sees
 ‘Patang sees me.’ (Tegey 1977, 169)

Vowel harmony also applies to preverbal clitics: In example (124), the high vowel in the verb triggers vowel harmony in the final syllable of the preverbal clitic *bāndi*.

- (124) wər= bāndi= (*bānde) xiçu
 it on step
 ‘We are stepping on it.’ (Tegey 1977, 227)

Summing up, the following facts can be established for the postlexical phonological process of vowel harmony (VH):

1. VH applies to all word categories if the phonological context is given.
2. Within the verbal complex, VH spreads to both groups of clitics.
3. VH cannot cross the boundary between two lexically stressed words (two individual prosodic words); i.e., vowel harmony is not restricted by the phonological phrase.
4. VH cannot spread to a 2P clitic that is outside of the verbal complex, even if it is directly preceding it.

As vowel harmony can spread to both groups of clitics given within the verbal complex, it can be assumed that the verbal complex itself forms one prosodic word, including the main verb, the clitics, and the negative particle (Figure 6.9).³⁶

³⁶If this prosodic word is in itself a phonological phrase, i.e., if there is a 1:1 relation between prosodic word and phonological phrase cannot be established here.

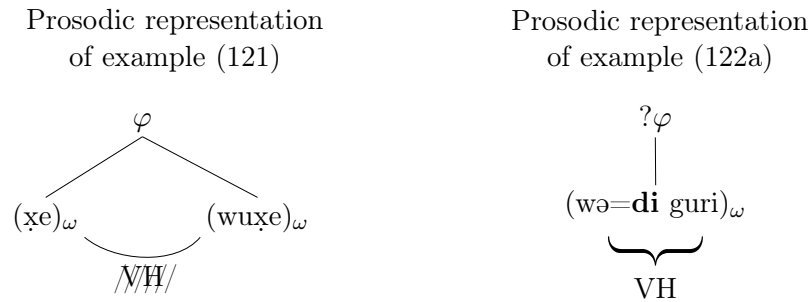


Figure 6.9: Prosodic representations and vowel harmony in examples (121) and (122a).

Thus, the following two postlexical phonological rules can be established so far:³⁷

- vowel coalescence: $\text{əa} \rightarrow \alpha / (\omega \text{ ?* } _ \text{ ?*})_\omega$
- vowel harmony (to apply repeatedly until fulfilled):

$$\left\{ \begin{array}{c} o \\ e \end{array} \right\} \rightarrow [+high] / (\omega \text{ ?* } _ \text{ C* } \left[\begin{array}{c} \text{V} \\ +high \end{array} \right] \text{ ?* })_\omega$$

A final question is the internal order of the postlexical phonological processes established so far: vowel harmony and vowel coalescence. As the following example shows, a strict linear ordering can be assumed.

- (125) $n\alpha$ =di xistə
 not.buy₁ should buy₂
 ‘He shouldn’t buy it.’ (Tegey 1977, 170)

A construction as in (125) can only be derived if one assumes that vowel coalescence applies first, followed by 2P clitic placement through prosodic inversion and finally followed by vowel harmony triggered by the second part of the verb and the now initial syllable with the high vowel [i] (Table 6.6). Had vowel coalescence not applied before, then the initial [a] would block any processes of vowel harmony.

³⁷As discussed in Chapter 4, Section 4.2.4.3, ? stands for any phoneme and * means that the preceding element can appear 0-∞ times. ?* consequently means that any phoneme can appear any number of times.

postlexical process	applied to
< input >	de nə axistə
↓	↓
vowel coalescence	de nα xistə
↓	↓
prosodic inversion	nα de xistə
↓	↓
vowel harmony	nα di xistə

Table 6.6: Postlexical phonological processes applied to example (125).

After having established that the outer shell of the verbal complex must be a prosodic word, the following phenomenon sheds some light on the internal grouping within the domain.

6.4.5.2 Initial /k/ deletion in complex predicates

As has been described in Section 6.2.3.3, Class III verbs are formed with a light verb and an adjective, adverb or noun. The light verb itself can occur in several shapes, where the choice of its form is an intersection of three categories: transitivity, tense, and aspect.³⁸ In the perfective, the stress falls on the initial lexical component. In the imperfective, on the other hand, stress falls on the light verb (or a suffix attached to it).

There are three possible realisations of the light verbs in the imperfective constructions: *-keg-* (non-past, intransitive), *-ked-* (past, intransitive) and *-kaw-* (transitive). The initial /k/ of these forms can be deleted depending on the final segment of the preceding lexical component: /k/ is retained if the final segment is a vowel ((126a)) and it is deleted if the preceding element ends in a consonant ((126b)).

- (126) a. asad ǰanəm wobə-**kaw**i
 Asad wheat water do
 ‘Asad was watering the wheat.’
- b. asad ǰanəm tit-Øaw*i* (*tit-**kaw**i)
 Asad wheat spread do
 ‘Asad was spreading the wheat.’ (Tegey 1977, 96)

In the perfective, where stress falls on the initial lexical component, this deletion never occurs with k-initial light verbs, which indicates some sort of prosodic border between the two elements.

- (127) dzhobəl **k-em**
 injured do
 ‘I injure...’ (Roberts 2000, 36, modified)

³⁸As the choice of the form is irrelevant for the current discussion, the light verbs are not further discussed here. For more information, see Tegey (1977, 95ff) and Tegey and Robson (1996, 105ff).

As Tegey and Robson (1996, 109) note, this relationship between lexical component and light verb is also reflected in writing. In the imperfective, the two words are written together; in the perfective, they are separated.

From this data it can be concluded that the occurrence of a perfective stress pattern comes along with a shift in prosodic constituency within the verb/the verbal complex. Otherwise, it would be difficult to justify why the initial /k/ can be deleted in the imperfective, but must be retained in the perfective construction.

As a consequence of these findings, Roberts (2000, 34) assumes the perfective class III construction to be phrased into two prosodic words in order to explain why the 2P clitic is positioned between the two elements. The imperfective, on the other hand, then forms only one prosodic word. While the latter notion can be argued for, it seems difficult to justify the perfective to form two prosodic words, especially if a solution to phrasing should extend to all verb classes: class I with the perfective marker and class II and class III with the stress shift. The conclusions on prosodic domains of postlexical phonological processes discussed in the previous sections add to this reluctance. The following section thus aims to establish a prosodic phrasing pattern that does justice to all of the facts presented above.

6.4.5.3 The prosodic phrasing of the verbal complex

As established in the previous section, vowel harmony is restricted to the domain of the prosodic word. As vowel harmony within the verb is *not* restricted by an aspect-caused stress shift, the consequence is that the perfective does not automatically form two prosodic words – otherwise, vowel harmony would be blocked within the perfective verb and across the verb boundary to preceding 2P clitics. Similarly vowel coalescence: Vowel coalescence does not occur between two prosodic words, but can be assumed to be restricted to one prosodic word as well. It would be difficult to explain then that the stressed element in the perfective and the remaining material form separate prosodic words if vowel coalescence and vowel harmony can occur between them.

On the other hand, lower prosodic units will not suffice either. The foot³⁹ as the crucial criterium does not explain why the initial /k/ of the auxiliary is deleted in the imperfective, but not in the perfective, as the underlying foot structure stays the same: It is only the stress that shifts from one foot to another.

The solution is a nested prosodic word structure that allows for (singular) boundaries weak enough for processes like vowel harmony to cross, but which can be strong enough to restrict processes like k-deletion on the other hand. These nested prosodic word structures⁴⁰ can be created in two ways:

³⁹As has been claimed by myself in Bögel (2010).

⁴⁰Such nested prosodic word structures have been claimed for similar combinations before, e.g., for host and clitics (Selkirk (1995), see also Chapter 4) and also in the context of affixation (see, e.g., Peperkamp (1997) and references therein).

1. **Each stressed item receives prosodic word status:** $(x\ x\ (\acute{x})_{\omega}\ x\ x)_{\omega}$

This approach would have to be restricted, in that it must not apply to the imperfective, but only to the perfective and the negative marker. In the imperfective, if it were only the light verb that would receive prosodic word status, then the initial k-deletion should again be blocked, as it cannot apply across the singular boundary between light verb and the preceding material. As this is not the case, the prosodic word status is either not granted in the imperfective (which is difficult to justify) or the status is applied differently:

2. **Each stressed item forms a prosodic word boundary to its right:**

$((x\ x\ \acute{x})_{\omega}\ x\ x)_{\omega}$

In this particular approach, the perfective, the imperfective and the negative marker are treated equally. In each case, a right prosodic word boundary is formed after the stressed element, grouping all previous elements into a first prosodic word while leaving all following elements to be part of the outer layer. If there is no material left, then the outer layer is superfluous.⁴¹

As the second possibility allows for phrasing without exceptions for the imperfective, the following analysis will thus assign the prosodic structure to the verbal complex as stated in (128):⁴²

(128) **Prosodic phrasing in the verbal complex:**

Assign a right prosodic word boundary after the stressed element.

The ‘landing site’ of the 2P clitics within the verbal complex can then be defined as follows:

(129) **Prosodic inversion in the verbal complex:** Within the verbal complex, a second position clitic is placed *after the first prosodic word*.

Note that this is of course only true if the 2P clitic is stranded in the preverbal position, i.e., if the enclitic is first in the intonational phrase and is consequently replaced to a position with an adequate host to its left.

In the following, this conclusion will be tested with a number of examples from the previous sections:

⁴¹Note that the two options also represent two major theoretical approaches to the prosody-syntax interface, albeit p-structure internally. The first option could be paralleled with MATCH and the second one with ALIGN (See Chapter 2 for an introduction).

⁴²If this concept also applies to the distribution of prosodic boundaries to elements outside of the verbal complex cannot be determined here, but awaits further research.

	construction	example	
1	$((w\acute{o})_{\omega} = \mathbf{di} \text{ guri})_{\omega}$	(122a)	after perfective prefix
2	$((\acute{t}\acute{e}l)_{\omega} = \mathbf{me} \text{ w\acute{o}h\acute{e}})_{\omega}$	(92b)	after stressed part of verb
3	$((n\acute{o})_{\omega} = \mathbf{ba} \text{ ga\acute{d}igi})_{\omega}$	(114)	after negative marker
4	$((r\alpha = \text{ta} = \text{pe} = \text{ga}\acute{n}\acute{d}\acute{o})_{\omega} = \mathbf{de})_{\omega}$	(87)	after verb and preverbal clitics
5	$((r\alpha = \text{ta} = \text{pe} = w\acute{o})_{\omega} = \mathbf{de} \text{ ga}\acute{n}\acute{d}\acute{o})_{\omega}$	(112)	after perfective prefix and Prev. Cl.

Table 6.7: Verification of the predictions made by (128) and (129).

As can be seen in Table 6.7, the 2P clitics are placed correctly after the perfective prefix ((1)) as well as after the first stressed element of a class II verb ((2)). In (3), the 2P clitic is placed after the negative marker, while (4) and (5) show the placement in imperfective and perfective constructions if preverbal clitics are present as well. It can be concluded that the prosodic approach proposed for the Pashto verbal complex and 2P clitic placement can be applied without exceptions.⁴³

6.4.5.4 The assignment of phrasing in the verbal complex

After having established the prosodic phrasing within the verbal complex, the question remains as to how this phrasing pattern is assigned: lexically or postlexically. This will be briefly discussed below.

First, a **postlexical phrasing** is assumed: In this case, the VC would project a prosodic word structure during the transfer of structure at the prosody-syntax interface. This prosodic word would then wrap the complete VC structure. In the following postlexical prosodic rephrasing, a right prosodic word boundary is inserted after each stressed element, where the first one can then be established as forming a host for the 2P clitic. This approach seems straightforward, but there is a problem: The insertion of the prosodic word boundary after the first stressed element is strictly postlexical, as the trigger is phonological (stress). Thus, this insertion cannot be restricted by reference to syntax as, e.g., ‘only insert boundaries within a prosodic word that has been formed by a VC’, as this would mean a breach of modularity. Consequently, this rule can apply to *all* prosodic words, that is, to nouns, adjectives and adverbs as well – and a justification for the creation of nested prosodic words out of each lexical element carrying lexical stress on the first syllable is not given.

An alternative would be **lexical phrasing**. In this approach, each lexically stressed element would be ‘prephrased’ within its p-form: The perfective verb would receive nested prosodic word status: $\Rightarrow ((\acute{v})_{\omega v})_{\omega}$, and the imperfective verb would receive simple prosodic word status. This approach makes the correct prediction as to the placement of the 2P clitics after the first prosodic word and it would also be justified to claim that the verb’s prosodic word structure stretches to include the

⁴³Note also that the tentative hypothesis proposed in Footnote 32 for the negative marker can be implemented as well with the phrasing proposed here. The underlying assumption would be that the negative marker is first if it carries the only stress in the construction, but is placed after the first prosodic word if there is another stressed element as well.

preverbal clitics: $\Rightarrow (\text{PreCL} = \text{v } \acute{v})_{\omega}$. However, if a negative marker is given, this negative marker receives independent prosodic word status as well and serves as a host to the 2P clitics. The expected phrasing would thus be: $\Rightarrow ((\text{nég}) = 2\text{P})_{\omega} (\text{v } \acute{v})_{\omega}$. This structure is problematic for postlexical phonological processes like vowel harmony which are confined by the domain of the prosodic word. As discussed above, vowel harmony would not be able to cross the boundaries from the prosodic word of the verb to the prosodic word of the negative marker which also incorporates the 2P clitics. On the other hand, it is not justified to argue for a ‘stretching’ of one prosodic word over the other to form a nested construction.

The only possible solution is a **mixed approach**. Lexical prosodic word status is assigned to all stressed elements, *except* for the verb. However, a *single* prosodic word boundary is assumed to the right of the verb-internal stress. The overall prosodic word structure is determined by the VC.

This leads to the following predictions: $\Rightarrow ((\text{PreCL} = \acute{v})_{\omega} = 2\text{P } \text{v})_{\omega}$ for a perfective construction and $\Rightarrow ((\text{PreCL} = \text{v } \acute{v})_{\omega} = 2\text{P})_{\omega}$ for an imperfective construction. Concerning the 2P clitics, this assumption leads to a correct placement, while at the same time allowing for the application of the phonological rules within the verbal complex and also for the integrity of the prosodic words formed by other lexical categories.

Proof for this assumption also comes from the interaction with the negative marker. In a perfective construction, the stressed negative marker is placed after the first stressed element of the prosodic word projected by VC: $\Rightarrow ((\acute{v})_{\omega} (\text{neg})_{\omega} \text{v})_{\omega}$. The reasons for this placement are unknown, but it would be much harder to explain why a separate prosodic word is placed behind another prosodic word boundary if the two elements did not share a common prosodic word domain.

The following conclusions about prosodic phrasing in the verbal complex can be drawn:

1. The lexical entry of the verb does not assign prosodic word status, but indicates the right prosodic word boundary after the stressed element. All other stressed elements have independent prosodic word status.
2. The corresponding left boundary is identical with the left boundary projected by VC.
3. Prosodic word status to the verbal complex is assigned during the *transfer of structure*.

After having established all the relevant facts on Pashto second position clitics, the following section focusses on an analysis with the LFG syntax-prosody interface as it was proposed in Chapter 5.

6.5 Pashto endoclitics: an LFG analysis

The following section demonstrates how the syntactic and prosodic/phonological findings and conclusions of the previous sections can be accounted for on the basis of the syntax-prosody interface developed in Chapter 5. The following endoclitic example, repeated from (96), will serve as a demonstration example.

- (130) $w\alpha$ =**ye** xla
 PERF.buy₁ it buy₂
 ‘(You) buy it.’ (Tegey 1977, 163)

Example (130) is a verb-initial perfective construction and as such part of the *prosodic* placement of 2P clitics. The a-initial verb *axla* ‘buy’ formally belongs to class I, as it marks the perfective aspect with the prefix *wə-*. Furthermore, (130) exhibits two postlexical phonological processes: vowel coalescence and prosodic inversion.

Before analysing c- and p-structure and the transfer processes between the two modules, the multidimensional lexical entries for the verb and the 2P clitic are established in Table 6.8.

s-form				p-form	
$w\alpha$ -axla	V	(↑ PRED)	= ‘axl<SUBJ, OBJ>’	P-FORM	[wáaxla]
		(↑ TENSE)	= past	SEGMENTS	/w ə a x l a/
		(↑ ASPECT)	= perf	METR. FRAME	‘σ) _ω σσ
		...			
ye	PRON	(↑ PRED)	= ‘ye’	P-FORM	[ye]
		(↑ PERS)	= 3	SEGMENTS	/y e/
		(↑ NUM)	= sg	METR. FRAME	=σ
		(↑ CL-TYPE)	= 2P		
		...			

Table 6.8: Lexical entries for example (130).

In Table 6.8, the verb is marked as transitive, subcategorizing for a subject and an object. Note that while the metrical frames for Degema (Chapter 5) assigned prosodic word status to the verb, this is not the case for Pashto for the reasons discussed above. Here, the verb only indicates a boundary following the stressed element. If (↑ ASPECT) = imperf and the stress is on the final syllable, this boundary would shift to the final edge of the verb in the corresponding metrical form.

The pronoun *ye* in the above table is marked as a prosodically deficient item (=σ). As it is also subject to syntactic replacement, the clitic is marked as a second position clitic within the s-form with the attribute-value pair (↑ CL-TYPE) = 2P. Two possible values exist for this particular attribute: 2P and PreCL (preverbal clitic). This distinction is important, as both groups display individual syntactic behavior: While the 2P clitics are syntactically placed after the first major constituent, the preverbal clitics are adjoined to VC. The clitics that are selected by the CCL node

thus have to be restricted by a constraint that demands the value of CL-TYPE to be 2P ((131)).⁴⁴

$$(131) \text{ CCL} \longrightarrow \dots (\text{PRON}) \dots \\ (\uparrow \text{CL-TYPE}) =_c \text{ 2P}$$

In Section 6.4.3, the following phrase structure rule was established for the syntactic placement of 2P clitics, where the first option determines the syntactic placement of the 2P clitics after the first major constituent, while the second option describes the case where the CCL is stranded clause-initially before the verbal complex.

$$(132) \text{ S} \longrightarrow \dots \\ \left[\begin{array}{l} \{ \text{XP CCL XP}^* \\ \text{CCL} \} \\ \text{VC} \end{array} \right] \\ \text{where XP} = \{ \text{NP} \mid \text{PP} \mid \text{AP} \mid \text{AdjP} \}$$

Together with the lexical entries proposed above, these phrase-structure rules lead to the following (simplified) c- and f-structure representations for (130).

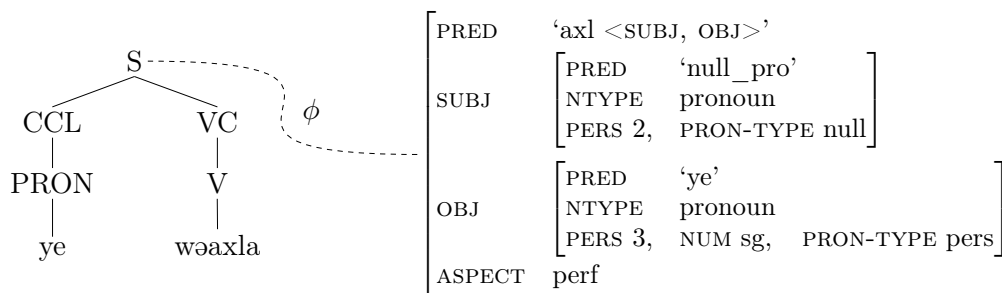


Figure 6.10: C- and f-structure representation of *wax ye xla* 'Buy it'.

In Figure 6.10, the f-structure representation shows the dropped subject argument ('null_pro'). The corresponding c-structure representation only includes CCL and VC as immediate daughters of S. The CCL node containing the 2P clitic *ye* is stranded in clause-initial position and consequently, the condition for prosodic 2P clitic placement is created.

From this syntactic starting position, the *transfer of structure* projects information on constituents to p-structure, while the s-forms of the terminal nodes are associated with their corresponding p-forms during the *transfer of vocabulary*. While CCL does not project any structural information to p-structure, VC projects a prosodic

⁴⁴As it is not entirely clear how the templatic order of the 2P clitics is derived, a full phrase-structural rule will not be developed at this point. It could well be that the order that surfaces in the p-string is determined solely by prosodic means.

word as it was discussed in Section 6.4.5.3. In order to avoid gaps in the prosodic hierarchy, it is also assumed that the prosodic word projected by VC is contained in a phonological phrase. As a consequence, the following projection annotation is added to the VC node:⁴⁵

$$\begin{array}{c} \text{VC} \\ (\natural(T(*)) S_{min} \text{ PHRASING}) = (\varphi(\omega \\ (\natural(T(*)) S_{max} \text{ PHRASING}) =)\omega)\varphi \end{array}$$

This annotation can be read as: ‘From all terminal nodes under VC, take the one with the minimum/maximum index in the p-diagram (i.e., the first/last syllable) and add the prosodic word/phonological phrase –boundary as a value to the attribute PHRASING’.⁴⁶ The clause itself, on the other hand, is matched to contain an intonational phrase.

$$\begin{array}{c} \text{S} \\ (\natural(T(*)) S_{min} \text{ PHRASING}) = (\iota \\ (\natural(T(*)) S_{max} \text{ PHRASING}) =)\iota \end{array}$$

This transfer of structural information and the associated *transfer of vocabulary*, which returns the information stored under the respective p-forms in Table 6.8, is illustrated in Figure 6.11.

Figure 6.11 shows the *transfer of structure* (\natural) on the one hand and the *transfer of vocabulary* (ρ) on the other hand. Information from both transfer processes is collected in the *preliminary p-diagram*, which is the input to p-structure. Note that no phonological processes have applied so far: The 2P clitic is still placed in the clause-initial position – but its prosodic deficiency has been exposed by accessing the respective p-form. The transfer process can thus be seen as a projection of syntactic and lexical information into another module; a ‘translation’ into phonological terms, where the syntactic information is exchanged for the corresponding phonological information. In the process, syntactic information is ‘forgotten’ and new, phonologically relevant information is discovered. After the transfer process is complete, further processing is secluded within the p-structure module (albeit ‘rechecking’ for information from the output of another module is possible, see Chapter 3, Section 3.5).

⁴⁵For reasons of simplification, the annotation of the phonological phrase will be excluded from Figure 6.11 below.

⁴⁶For a detailed introduction into the transfer of structure at the syntax-prosody interface, see Chapter 5, Section 5.4.2.

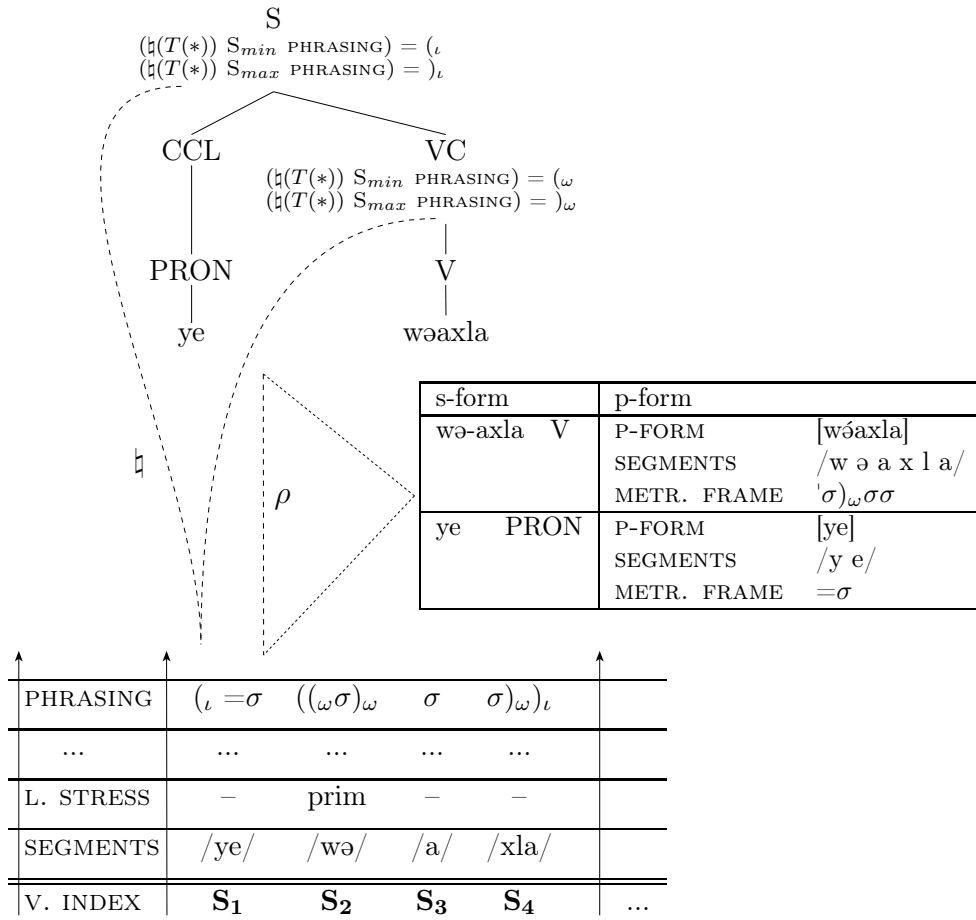


Figure 6.11: Transfer of structure and vocabulary: *ye wəaxle* ‘Buy it’.

Once the information from the syntax and the lexicon has been transferred, the preliminary p-diagram is subjected to a cascade of postlexical phonological rules and constraints within p-structure. For example (130), two postlexical phonological processes apply: vowel coalescence and prosodic inversion of the 2P clitic. As demonstrated in Table 6.6, vowel coalescence must apply *before* 2P clitic placement; otherwise, the output for (130) would be **wə ye axla*.

1. Application of vowel coalescence to the preliminary p-diagram generated in Figure 6.11: $\text{əa} \rightarrow \alpha / (\omega \text{ ?}^* _ \text{ ?}^*) \omega$

↑	↑	↑	↑
PHRASING	$(\iota = \sigma$	$((\omega \sigma)_\omega$	$\sigma)_\omega)_\iota$
...
L. STRESS	–	prim	–
SEGMENTS	/ye/	/wα/	/xla/
V. INDEX	S₁	S₂	S₃

If /ə/ and /a/ are directly adjacent within the domain of a prosodic word, then vowel coalescence applies.

2. Prosodic inversion is applied to the intermediate p-diagram generated in (1.):
 $(\iota = \sigma + (\sigma =)^* \omega \rightarrow (\iota (\sigma =)^* \omega = \sigma +$

↑	↑	↑	↑
PHRASING	$(\iota ((\omega \sigma)_\omega$	$= \sigma$	$\sigma)_\omega)_\iota$
...
L. STRESS	prim	–	–
SEGMENTS	/wα/	/ye/	/xla/
V. INDEX	S₁	S₂	S₃

Any number of enclitics that are placed in the initial position of the intonational phrase (possibly followed by proclitics) are placed after the next prosodic word.

3. The resulting p-string is $w\alpha \text{ ye } xla$, as given in (130).

The processes described in this section from the perspective of production are completely reversible. That is, given an input like the p-string in (3.), the analysis can then be reversed step by step to the syntactic analysis given in Figure 6.10.

6.6 Summary and conclusion

This chapter has proposed a new approach to Pashto 2P en(do)clisis. Two distinct processes of second position placement have been established: First, syntactic placement of 2P clitics after the first major constituent preceding the verbal complex and second, prosodic inversion of 2P clitics to the position after the first prosodic word if the clitic is stranded in the clause-initial (intonational-phrase initial) position. These claims have been supported by a number of syntactic and phonological facts:

- a. It has been established that the major constituent hosting the 2P clitic during syntactic placement does not have to be constrained to carry stress, because

any constituent that does not carry stress syntactically becomes a part of the verbal complex. Thus, the attempt to find a prosodic domain that can account for large major constituents on the one hand and partials of the verbs on the other hand is irrelevant. Syntactically, the 2P clitics will always precede the verbal complex no matter if destressed constituents have been attached to VC or not. The placement of 2P clitics within or after the verb is always a matter of postlexical prosodic inversion.

- b. Several phonological processes shed light on the external and internal structure of the verbal complex. It has been established that the VC node projects a prosodic word to p-structure. Furthermore, it was assumed that the verb does not project a complete prosodic word by itself, but that it indicates a prosodic word boundary after the stressed element. This results in a nested prosodic word structure within the verbal complex, which allows for processes like vowel coalescence and vowel harmony to apply internally, but restricts processes like initial k-deletion to constructions where no singular boundary is given

The chapter concluded with an analysis of these findings via the syntax–prosody interface as it has been established in this thesis. It was shown that the architecture developed here allows for a straightforward analysis of a complex phenomenon like second position en(do)clisis in Pashto by assuming two distinct modules, syntax and p-structure. The p-structure module mostly reflects the linear order as provided by the c-structure, but allows for an intervention in cases of replacement on the basis of prosodic constraints.

Chapter 7

Summary and conclusion

This thesis introduced a new approach to the syntax–prosody interface in LFG and extended the overall correspondence architecture to include a p-structure module that allows for a full analysis of postlexical phonological phenomena.

As discussed in Chapter 2, language is assumed to be generated between the two notions of FORM and MEANING by means of a number of modules, each with its own primitives and principles, its own ‘vocabulary’. Modules are not assumed to be fully encapsulated units, but build up in ‘parallel’: i.e., once a module begins to build up structure, fragmental information becomes available to other modules which in turn begin to build up structure in parallel. This information at the interfaces is transferred via LFG’s projection functions that ‘translate’ information into the module’s ‘native language’.

Concrete language phenomena are thus either analysed from the perspective of *comprehension* or from the perspective of *production*, without referring to actual performance factors. Rather, an abstract ‘road-map’, a guidance for the analysis in a case of real performance is developed for each phenomenon.

The relation between the syntactic and the phonological module is determined by two transfer processes: the *transfer of vocabulary* and the *transfer of structure*. The *transfer of vocabulary* assumes a multidimensional lexicon, which mediates between the syntactic and the phonological form of each lexical entry, thus transferring information on a ‘lower’ level. The *transfer of structure*, on the other hand, assumes a direct relation between the two modules where information on structural data is exchanged in form of prosodic constituency. Note, however, as the evaluation of different approaches to the syntax–prosody interface in Chapter 2 showed, there is no perfect mapping between the two modules. Therefore, the mapping of syntactic structure to prosodic structure is considered to be preliminary; the default assumption is that prosodic restructuring applies to this preliminary input during production, as it was demonstrated in Chapter 4, 5 and 6. That is, non-isomorphism is not considered to be exceptional, but is rather the default.

One important conclusion with respect to the directional analysis is the fact

that an analysis from a specific perspective is not necessarily reversible, which is mainly due to the fact that while syntactic structure influences prosodic structure to a considerable extent, the reverse cannot be said to be true. As was discussed in Chapter 3, prosodic structure can only occasionally determine syntactic structure (e.g., in the case of ambiguities). In general, however, prosodic structure is not ‘restructured’ to match syntactic structure during comprehension. This was briefly discussed with reference to Degema in Chapter 5, Section 5.4.5, but references to the topic can also be found in other chapters.

The p-structure module has been described by means of a fine-grained and compact representation that reflects the linear and complex nature of speech on the one hand and allows for a representation of (hierarchical) prosodic constituency on the other hand: The p-diagram (Chapter 3). Its syllable-based foundation enables a representation independently of the (rather instable) formation of the prosodic hierarchy and is, in principle, theory-independent, as it is left to the researcher which data is encoded and how the data is interpreted.

In addition to the p-diagram, a set of postlexical phonological rules are assumed to apply within p-structure on the basis of the information encoded in the p-diagram (Chapters 4 – 6). Thus, a wide range of postlexical phonological phenomena and prosodic restructuring can be accounted for. In the case of production, these postlexical phonological rules apply to the ‘preliminary’ p-diagram as it was constructed by the two transfer processes at the interface. The resulting output, the ‘final p-diagram’, is the contribution of syntax, lexicon and postlexical phonology to the speech signal; the linear order of its elements is the surface (p-)string. P-structure is thus more than the mere representation of prosodic constituency, but allows for a full analysis of the phonological aspects of language, thus considerably expanding the range of phenomena that can be analysed with LFG.

One class of phenomena discussed in this thesis were clitics that change their linear position because of specific prosodic requirements, either as endoclitics (Chapter 5) or as second position clitics that can surface as endoclitics as well (Chapter 6). These clitics are problematic, as the interference of prosodic constraints with syntactic linear order and morphological integrity seemingly violates the principle of lexical integrity as well as modularity. However, the interface architecture as it was developed in this thesis allows for a straightforward analysis by assuming a non-parallelism between the p-string and the s-string. The underlying assumption is that the s-string represents syntactic linear order as it was determined by c- and f-structure, but that the p-string represents the linear order of the elements as determined by prosodic constraints. While p- and s-string are mostly parallel, the p-string can differ from the s-string in cases where prosodic constraints require the postlexical reordering of specific elements. The underlying architecture is represented in Figure 7.1.

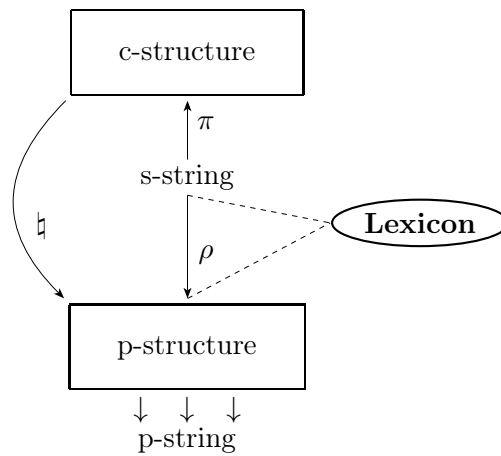


Figure 7.1: The prosody-syntax interface in LFG (final version).

These architectural assumptions allow for an analysis of endoclitisis in the p-structure module independently from syntax. As a result, the modularity of the different components is preserved and the concept of lexical integrity is kept intact.

As mentioned in the introduction, this thesis contains a wide range of open questions for future research. One interesting aspect is the understanding of this thesis as a blueprint for a computational implementation, which would allow for the ParGram grammars to include (and compare) aspects of phonology as well. Another more theoretical direction would be the testing of the current interface against possibly challenging phenomena as discussed by Zec and Inkelas (1990) in Chapter 2, Section 2.5.1 (see also Seidl (2001) or Elordieta (1997)). And last, but not least, another far-reaching topic is the relation between the syntax-prosody interface as it is presented in this thesis and other modules that can influence prosodic structuring as well, for example, information structure.

Appendix

Stimuli for the production experiment discussed in Chapter 3:

1. A paragraph to get the speaker's baseline and to include some 'hidden' (thus more spontaneous) constructions.

#####

Als der Prozess endlich begann und die Presse endlich ihre Plätze eingenommen hatte, war noch lange nicht klar, ob der Gärtner, der Kellner oder die Dame für *den Tod des Dieners* verantwortlich waren. Nachdem das Gericht eine Zusammenfassung der Fakten gegeben hatte, begab sich der Gärtner in den Zeugenstand und wurde hier durch *den Anwalt der Dame* befragt. Der Gärtner schien jedoch ein lückenloses Alibi zu haben. Daraufhin musste die Dame in den Zeugenstand und wurde *vom Anwalt des Kellners* befragt. Die Fragen die *der Anwalt der Dame* stellte, konnten von ihr nicht zufriedenstellend beantwortet werden, was den Verdacht auf die Dame lenkte.

Doch als sich die Jury anschließend traf und sie den ganzen Prozess noch einmal durchsprachen, erinnerten sie sich an eine Begebenheit: Der Richter hatte der Dame eine Frage gestellt, die sie etwas verwirrend beantwortet hatte. Daraufhin hatte dem Richter doch noch *der Anwalt der Dame* geantwortet, um Missverständnisse auszuschließen. Die Jury war sich daraufhin einig: Mit dieser Antwort hatte *der Anwalt der Dame* geholfen, sich von dem Verdachtsmoment zu befreien.

#####

2. Unambiguous sentences (with masculine DP)

Two noun phrases, trochaic, disyllabic, masculine; each subject saw only one sentence of each pair.

⇒ **Set A**

- 1_{Dat} Keiner war überrascht, dass **der Schwager dem Jäger** zustimmte.
- 2_{Gen} Der Richter sah erstaunt auf, als **der Anwalt des Lehrers** widersprach.
- 3_{Dat} Alle schwiegen, als **der Schneider dem Anwalt** antwortete.
- 4_{Gen} Die Eltern freuten sich, dass **der Onkel des Sängers** gratulierte.
- 5_{Dat} Jeder wusste, dass **der Enkel dem Opa** fehlte.
- 6_{Gen} Es war klar, dass **der Erbe des Königs** drohte.

⇒ **Set B**

- 1_{Gen} Keiner war überrascht, dass **der Schwager des Jägers** zustimmte.
- 2_{Dat} Der Richter sah erstaunt auf, als **der Anwalt dem Lehrer** widersprach.
- 3_{Gen} Alle schwiegen, als **der Schneider des Anwalts** antwortete.
- 4_{Dat} Die Eltern freuten sich, dass **der Onkel dem Sänger** gratulierte.
- 5_{Gen} Jeder wusste, dass **der Enkel des Opas** fehlte.
- 6_{Dat} Es war klar, dass **der Erbe dem König** drohte.

3. Ambiguous sentences (with feminine DP)

Two noun phrases, the first is masculine, the second feminine, followed by an intransitive verb with optional dative object. All subjects have to speak all sentences. To provide a disambiguation for the speaker, the context is given.

1. Das Gericht war sehr überrascht, als **der Anwalt der Diva** widersprach.
2. Um alles mitzubekommen, musste **der Fahrer der Dame** zuhören.
3. Alle freuten sich, als **der Onkel der Nonne** gratulierte.
4. Um rechtzeitig fertig zu werden, musste **der Schwager der Tante** helfen.
5. Die Enkel waren daher überrascht, als **der Gärtner der Oma** zustimmte.
6. Jeder bemerkte, dass **der Partner der Freundin** fehlte.
7. Keiner dachte sich etwas dabei, als **der Diener der Gräfin** folgte.
8. Alle hörten gespannt zu, als **der Lehrer der Schwäbin** antwortete.
9. Die Anwesenden waren sehr überrascht, dass **der Rabe der Heldin** gehorchte.

Bibliography

- Abercrombie, David. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- Allbritton, David W., McKoon, Gail and Ratcliff, Roger. 1996. Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(3), 714–735.
- Alsina, Alex. 2010. The Catalan definite article as lexical sharing. In *Proceedings of LFG10*, Stanford, CA: CSLI Publications.
- Anderson, Stephen. 2005. *Aspects of the Theory of Clitics*. Oxford [a.o.]: Oxford University Press.
- Andréasson, Maia. 2007. The architecture of i-structure. In *Proceedings of LFG07*, CSLI Publications.
- Asudeh, Ash. 2001. Linking, optionality, and ambiguity in Marathi. In P. Sells (ed.), *Formal and Empirical Issues in Optimality Theoretic Syntax*, pages 257–312, CSLI Publications.
- Asudeh, Ash. 2006. Direct compositionality and the architecture of LFG. In M. Butt, M. Dalrymple and T. H. King (eds.), *Intelligent Linguistic Architectures: Variations on themes by Ronald M. Kaplan*, Stanford, CA: CSLI Publications.
- Asudeh, Ash. 2009. Adjacency and locality: a constraint-based analysis of complementizer-adjacent extraction. In *Proceedings of LFG09*, CSLI Publications.
- Asudeh, Ash, Dalrymple, Mary and Toivonen, Ida. 2013. Constructions with lexical integrity. *Journal of Language Modelling* 1(1), 1–54.
- Asudeh, Ash and Toivonen, Ida. 2010. Lexical-Functional Grammar. In B. Heine and H. Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, Chapter 17, pages 425–458, Oxford: Oxford University Press.
- Austin, Peter K. and Bresnan, Joan. 1996. Non-configurationality in Australian Aboriginal languages. *Natural Language and Linguistic Theory* 14(2), 215–268.
- Babrakzai, Farooq. 1999. *Topics in Pashto Syntax*. Ph.D.thesis, University of Michigan, Ann Arbor.
- Bader, Markus. 1996. *Sprachverstehen – Syntax und Prosodie beim Lesen*. Opladen: Westdeutscher Verlag.
- Baker, Peter S. 2007. *Introduction to Old English*. Malden, MA and Oxford and Carlton Victoria: Blackwell, second edition edition.

- Beckman, Mary and Ayers Elam, Gayle. 1997. Guidelines for ToBI labelling, available at http://www.ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdf [retrieved 2.3.2015].
- Beckman, Mary E. and Hirschberg, Julia. 1994. The ToBI Annotation Conventions, available at http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html [retrieved 12.3.2015].
- Beckman, Mary E., Hirschberg, Julia and Shattuck-Hufnagel, Stefanie. 2005. The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, Chapter 2, pages 9–54, Oxford University Press.
- Beckman, Mary E. and Pierrehumbert, Janet B. 1986. Intonational structure in English and Japanese. *Phonology Yearbook* 3, 255–309.
- Beesley, Kenneth R. and Karttunen, Lauri. 2003. *Finite State Morphology*. Stanford, CA: CSLI Publications.
- Bock, Kathryn and Levelt, Willem J.M. 1994. Language production: grammatical encoding. In M. A. Gernsbacher (ed.), *Handbook of Psycholinguistics*, Chapter 29, pages 945–984, London: Academic Press.
- Boersma, Paul and Weenink, David. 2013. Praat: doing phonetics by computer [Computer program, Version 5.3.56], available at <http://www.praat.org/> [retrieved 15.09.2013].
- Bögel, Tina. 2010. Pashto (endo)clitics in a parallel architecture. In *Proceedings of LFG10*, Stanford, CA: CSLI Publications.
- Bögel, Tina. 2012. The p-diagram – a syllable-based approach to p-structure. In *Proceedings of LFG12*, Stanford, CA: CSLI Publications.
- Bögel, Tina. 2013. A prosodic resolution of German case ambiguities. In *Proceedings of LFG13*, Stanford, CA: CSLI Publications.
- Bögel, Tina. 2014. Degema and the string interface. In *Proceedings of LFG14*, Stanford, CA: CSLI Publications.
- Bögel, Tina and Butt, Miriam. 2013. Possessive clitics and ezafe in Urdu. In K. Börjars, D. Denison and A. K. Scott (eds.), *Morphosyntactic Categories and the Expression of Possession*, pages 291–322, John Benjamins.
- Bögel, Tina, Butt, Miriam, Kaplan, Ronald M., King, Tracy H. and III., John T. Maxwell. 2010. Second position and the prosody-syntax interface. In *Proceedings of LFG10*, Stanford, CA: CSLI Publications.

- Bögel, Tina, Butt, Miriam, Kaplan, Ronald M., King, Tracy Holloway and Maxwell III., John T. 2009. Prosodic phonology in LFG: a new proposal. In *Proceedings of LFG09*, CSLI Publications.
- Bögel, Tina, Butt, Miriam and Sulger, Sebastian. 2008. Urdu Ezafe and the Morphology-Syntax Interface. In *Proceedings of LFG08*, Stanford, CA: CSLI Publications.
- Bohnacker, Ute. 2013. Null subjects in Swabian. *Studia Linguistica* 67(3), 257–289.
- Booij, Geert. 1988. On the relation between lexical and prosodic phonology. In P. M. Bertinetto and M. Loporcaro (eds.), *Certamen phonologicum. Papers from the 1987 Cortona Phonology meeting*, pages 63–75, Rosenberg und Sellier.
- Bresnan, Joan. 1978. A realistic transformational grammar. In M. Halle, J. Bresnan and G. A. Miller (eds.), *Linguistic Theory and Psychological Reality*, pages 1–59, Cambridge, MA: MIT Press.
- Bresnan, Joan. 2000. Optimal syntax. In J. Dekkers, F. van der Leeuw and J. van de Weijer (eds.), *Optimality Theory: Phonology, Syntax, and Acquisition*, pages 334–385, Oxford: Oxford University Press.
- Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Malden, MA: Blackwell.
- Bresnan, Joan and Kaplan, Ronald M. 1982. Lexical-Functional Grammar: A Formal System for Grammatical Representation. In J. Bresnan (ed.), *The Mental Representation of Grammatical Relations*, pages 173–281, Cambridge, MA: MIT Press.
- Bresnan, Joan and Mchombo, Sam. 1995. The Lexical Integrity Principle: Evidence from Bantu. *Natural language and Linguistic Theory* 13(2), 181–254.
- Broadwell, George A. 2008. Turkish Suspended Affixation is Lexical Sharing. In *Proceedings of LFG08*, CSLI Publications.
- Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Dissertations in Linguistics, Stanford, CA: CSLI Publications.
- Butt, Miriam, Bögel, Tina, Hautli, Annette, Sulger, Sebastian and Ahmed, Tafseer. 2012. Identifying Urdu complex predication via bigram extraction. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, pages 409–424, Mumbai, India.
- Butt, Miriam, Dyvik, Helge, King, Tracy H., Masuichi, Hiroshi and Rohrer, Christian. 2002. The Parallel Grammar Project. In *Proceedings of the COLING-2002 Workshop on Grammar Engineering and Evaluation*.

- Butt, Miriam and Kaplan, Ronald M. 2002. 'The Morphology-Syntax Interface', Handout presented at the Lexical Functional Grammar Conference, Athens, Greece.
- Butt, Miriam and King, Tracy Holloway. 1997. Null elements in discourse structure, unpublished manuscript.
- Butt, Miriam and King, Tracy Holloway. 1998. Interfacing Phonology with LFG. In *Proceedings of LFG98*, Stanford, CA: CSLI Publications.
- Chen, Matthew Y. 1987. The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4, 109–149.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, Noam. 2001. Derivation by phase. In M. Kenstowicz (ed.), *Ken Hale: A Life in Language*, pages 1–52, Cambridge, MA: MIT Press.
- Chomsky, Noam. 2006. *Language and Mind*. New York: Cambridge University Press, third edition.
- Chomsky, Noam and Halle, Morris. 1968. *The Sound Pattern of English*. New York: Harper and Row.
- Crouch, Richard, Dalrymple, Mary, Kaplan, Ronald M., King, Tracy H., Maxwell III, John T. and Newman, Paula. 2015. *XLE documentation*. Palo Alto, CA: Palo Alto Research Center, online documentation.
- Crowhurst, Megan J. 1998. *Um* infixation and prefixation in Toba Batak. *Language* 74(3), 590–604.
- Dalrymple, Mary. 2001. *Lexical Functional Grammar*. San Diego [a.o.]: Academic Press.
- Dalrymple, Mary and Mycock, Louise. 2011. The prosody-semantics interface. In *Proceedings of LFG11*, Stanford, CA: CSLI Publications.
- Dalrymple, Mary and Nikolaeva, Irina. 2011. *Objects and information structure*. Cambridge [a.o.]: Cambridge University Press.
- Dehé, Nicole. 2014a. Final devoicing of /l/ in Reykjavík Icelandic. In *Proceedings of the 7th International Conference on Speech Prosody*, Dublin, Ireland.
- Dehé, Nicole. 2014b. *Parentheticals in Spoken English: the Syntax-Prosody Relation*. Cambridge [a.o.]: Cambridge University Press.

- Dehé, Nicole. 2015. Particle verbs in Germanic. In P. O. Müller, I. Ohnheiser, S. Olsen and F. Rainer (eds.), *Word Formation. An International Handbook of the Languages of Europe*, pages 611–626, Mouton de Gruyter.
- Dell, Gary S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93(3), 283–321.
- Di Sciullo, Anne-Marie and Williams, Edwin. 1987. *On the Definition of Word*. Cambridge, MA [a.o.]: MIT Press.
- Dipper, Stefanie. 2003. *Implementing and Documenting Large-Scale Grammars – German LFG*. Ph.D.thesis, Institut für maschinelle Sprachverarbeitung (IMS), University of Stuttgart.
- Dost, Ascander H. 2007. *Linearization, Square Pegs, and Round Holes*. Ph.D.thesis, University of California, Santa Cruz.
- Eisenberg, Peter. 2006. *Grundriss der deutschen Grammatik: Das Wort*. Stuttgart; Weimar: Metzler, third edition.
- Elordieta, Gorka. 1997. *Morphosyntactic Feature Chains and Phonological Domains*. Ph.D.thesis, University of Southern California.
- Elordieta, Gorka. 2008. An overview of theories of the syntax-phonology interface. *ASJU* 42(1), 209–286.
- Fodor, Jerry A. 1983. *The Modularity of Mind*. Cambridge, MA [a.o.]: MIT Press.
- Fodor, Jerry A., Bever, Thomas G. and Garrett, Merrill F. 1974. *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar*. New York [a.o.]: McGraw-Hill.
- Forst, Martin, King, Tracy H. and Laczkó, Tibor. 2010. Particle verbs in computational LFGs: issues from English, German, and Hungarian. In *Proceedings of LFG10*, Stanford, CA: CSLI Publications.
- Frank, Anette, King, Tracy Holloway, Kuhn, Jonas and Maxwell III, John T. 1998. Optimality Theory style constraint ranking in large-scale LFG grammars. In *Proceedings of LFG98*, Stanford, CA: CSLI Publications.
- Franks, Steven and King, Tracy Holloway. 2000. *A Handbook of Slavic Clitics*. New York: Oxford University Press.
- French, Koleem Matsuda. 1988. *Insights into Tagalog Reduplication, Infixation and Stress from Non-linear Phonology*. Masters Thesis, University of Texas, Arlington.

- Friederici, Angela, Steinhauer, Karsten and Frisch, Stefan. 1999. Lexical integration: Sequential effects of syntactic and semantic information. *Memory & Cognition* 27(3), 438–453.
- Frota, Sónia. 2000. *Prosody and Focus in European Portuguese*. New York: Garland.
- Frota, Sónia. 2012. Prosodic structure, constituents, and their implementation. In A. C. Cohn, C. Fougeron and M. K. Huffman (eds.), *The Oxford Handbook of Laboratory Phonology*, Chapter 11, pages 255–265, Oxford University Press.
- Fudge, Erik. 1999. Words and feet. *Journal of Linguistics* 35(2), 273–96.
- Fudge, Erik C. 1969. Syllables. *Journal of Linguistics* 5(2), 253–286.
- Fulop, Sean, Kari, Ethelbert E. and Ladefoged, Peter. 1998. An acoustic study of the tongue root contrast in Degema vowels. *Phonetica* 55, 80–98.
- Gerlach, Birgit and Grijzenhout, Janet. 2000. Clitics from different perspectives (Introduction). In B. Gerlach and J. Grijzenhout (eds.), *Clitics in Phonology, Morphology and Syntax*, pages 1–29, Amsterdam: John Benjamins.
- Ghini, Mirco. 1993. Phi-formation in Italian: a new proposal. *Toronto Working Papers in Linguistics* 12(2), 41–78.
- Goedemans, Rob and van der Hulst, Harry. 2013. Rhythm types. In M. S. Dryer and M. Haspelmath (eds.), *The World Atlas of Language Structures Online*, Chapter 17, Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Goldsmith, John. 1976. *Autosegmental Phonology*. Ph.D.thesis, Massachusetts Institute of Technology.
- Goldsmith, John. 2011. The syllable. In J. Goldsmith and J. Riggle (eds.), *The Handbook of Phonological Theory*, Chapter 6, pages 164–196, Malden, MA: Wiley-Blackwell, second edition.
- Gollrad, Anja, Sommerfeld, Esther and Kügler, Frank. 2010. Prosodic cue weighting in disambiguation: Case ambiguity in German. In *Proceedings of Speech Prosody*, Chicago.
- Grice, Martine and Baumann, Stefan. 2002. Deutsche intonation und GToBI. *Linguistische Berichte* 191, 267–298.
- Grijzenhout, Janet and Kabak, Barış. 2009. Prosodic phonology: an appraisal. In *Phonological Domains: Universals and Deviations*, pages 1–11, Mouton de Gruyter.
- Gussenhoven, Carlos. 1991. The English Rhythm Rule as an Accent Deletion Rule. *Phonology* 8(1), 1–35.

- Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Haag-Merz, Christine. 1996. *Pronomen im Schwäbischen: Syntax und Erwerb*. Marburg: Tectum.
- Hall, Tracy Alan. 1999. Phonotactics and the prosodic structure of German function words. In T. A. Hall and U. Kleinhenz (eds.), *Studies on the Phonological Word*, pages 99–132, John Benjamins.
- Hall, Tracy Alan. 2011. *Phonologie – Eine Einführung*. Berlin: de Gruyter, second edition.
- Halle, Morris. 1973. Prolegomena to a theory of word formation. *Linguistics Inquiry* 4(1), 3–16.
- Halpern, Aaron L. 1995. *On the Placement and Morphology of Clitics*. Stanford, CA: CSLI Publications.
- Halpern, Aaron L. 1996. Introduction. In A. L. Halpern and A. M. Zwicky (eds.), *Approaching Second: Second Position Clitics and Related Phenomena*, Stanford, CA: CSLI Publications.
- Harris, Alice C. 2002. *Endoclitics and the Origins of Udi Morphosyntax*. Oxford [a.o.]: Oxford University Press.
- Hayes, Bruce. 1989. The prosodic hierarchy in meter. In P. Kiparsky and G. Youmans (eds.), *Rhythm and Meter*, pages 201–260, Orlando, FL: Academic Press.
- Hayes, Bruce. 2009. *Introductory Phonology*. Chichester: Wiley-Blackwell.
- Hayes, Bruce and Lahiri, Aditi. 1991. Bengali Intonational Phonology. *Natural Language and Linguistic Theory* 9(1), 47–96.
- Heinz, Jeffrey, Kobele, Gregory M. and Riggle, Jason. 2009. Evaluating the complexity of Optimality Theory. *Linguistic Inquiry* 40(2), 277–288.
- Hock, Hans H. 1996. Who's on first? Towards a prosodic account of P2 clitics. In A. L. Halpern and A. M. Zwicky (eds.), *Approaching Second: Second Position Clitics and Related Phenomena*, Chapter 6, pages 199–270, Stanford, CA: CSLI Publications.
- Hyman, Larry M. 2006. Affixation by place of articulation: rare and mysterious. *UC Berkeley Phonology Lab Annual Report*, University of California, Berkeley.
- Idsardi, William J. 2006. A simple proof that Optimality Theory is computationally intractable. *Linguistic Inquiry* 37(2), 271–275.

- Inkelas, Sharon. 1990. *Prosodic Constituency in the Lexicon*. Outstanding Dissertations in Linguistics, New York: Garland.
- Inkelas, Sharon and Zec, Draga. 1995. Syntax-phonology interface. In J. Goldsmith (ed.), *The handbook of phonological theory*, Chapter 15, pages 535–549, Cambridge, MA: Blackwell.
- Ito, Junko and Mester, Armin. 1992. Weak layering and word binarity. In *LRC Working Paper*, volume 92, Santa Cruz, CA: UC Santa Cruz.
- Jackendoff, Ray. 2002. *Foundations of Language*. New York: Oxford University Press.
- Jackendoff, Ray. 2010. The Parallel Architecture and its place in cognitive science. In B. Heine and H. Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*, Chapter 23, pages 583–605, Oxford: Oxford University Press.
- Jenewari, Charles E. W. 1980. *A Linguistic Guide to Spoken Kalabari*. Port Harcourt: School of Humanities, University of Port Harcourt.
- Jurafsky, Daniel and Martin, James H. 2009. *Speech and Language Processing*. Upper Saddle River, NJ: Prentice Hall, Pearson Education International, second edition.
- Kabak, Barış. 2007. Turkish suspended affixation. *Linguistics* 45(2), 311–347.
- Kahn, Daniel. 1968. *Syllable-based generalizations in English phonology*. Ph. D.thesis, Massachusetts Institute of Technology.
- Kahn, Daniel. 1976. *Syllable-based Generalizations in English Phonology*. Ph. D.thesis, Massachusetts Institute of Technology.
- Kaisse, Ellen M. 1981. Separating phonology from syntax: A reanalysis of Pashto cliticization. *Journal of Linguistics* 17, 197–208.
- Kaisse, Ellen M. 1985. *Connected Speech: The Interaction of Syntax and Phonology*. Orlando: Academic Press.
- Kaplan, Ronald M. 1987. Three seductions of computational psycholinguistics. In P. Whitelock, M. McGee Wood, H. L. Somers, R. Johnson and P. Bennett (eds.), *Linguistic Theory and Computer Applications*, pages 149–181, London [a.o.]: Academic Press.
- Kaplan, Ronald M. 1995. The formal architecture of Lexical-Functional Grammar. In M. Dalrymple, R. M. Kaplan, J. T. Maxwell III and A. Zaenen (eds.), *Formal Issues in Lexical-Functional Grammar*, pages 7–27, Stanford: CSLI Publications.
- Kari, Ethelbert E. 2002. On endoclitics: some facts from Degema. *Journal of Asian and African Studies* 63.

- Kari, Ethelbert E. 2003. *Clitics in Degema: A Meeting Point of Phonology, Morphology, and Syntax*. Ph.D.thesis, Research Institute for Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies, Tokyo.
- Kari, Ethelbert E. 2004. *A Reference Grammar of Degema*. Köln: Köppe.
- Kari, Ethelbert E. 2007. Vowel harmony in Degema, Nigeria. *African Study Monographs* 28(2), 87–97.
- Kari, Ethelbert E. 2012. Endoclitization and the lexical integrity hypothesis: Insights from Degema. In M. P. Larrañaga and P. Guijarro-Fuentes (eds.), *Pronouns and Clitics in Early Language*, Berlin: De Gruyter Mouton.
- Karttunen, Lauri. 1998. The proper treatment of optimality in computational phonology. In *Proceedings of the International Workshop on Finite-State Methods in Natural Language Processin, FSMNLP'98*.
- Kenstowicz, Michael J. and Kisseberth, Charles W. 1977. *Topics in Phonological Theory*. New York: Academic Press.
- King, Tracy H. 2005. Cliticizing LFG. In *Proceedings of LFG05*, Stanford, CA: CSLI Publications.
- King, Tracy Holloway. 1997. Focus domains and information structure. In *Proceedings of LFG97*, CSLI Publications.
- Kingston, John. 2007. The phonetics-phonology interface. In P. de Lacy (ed.), *The Cambridge Handbook of Phonology*, Chapter 17, pages 401–434, Cambridge [a.o.]: Cambridge University Press.
- Kiparsky, Paul. 1982a. From cyclic phonology to lexical phonology. In H. van der Hulst and N. Smith (eds.), *The structure of phonological representations*, volume 1, pages 131–175, Dordrecht: Foris.
- Kiparsky, Paul. 1982b. Lexical morphology and phonology. In *Linguistics in the Morning Calm*, pages 3–91, Seoul, Korea: Hanshin Publishing Company.
- Kleinhenz, Ursula. 1998. *On Words and Phrases in Phonology*. Ph.D.thesis, University of Tübingen.
- Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3–4), 243–276.
- Kuhn, Jonas. 2001. Generation and parsing in Optimality Theoretic syntax: issues in the formalization of OT-LFG. In P. Sells (ed.), *Formal and Empirical Issues in Optimality Theoretic Syntax*, pages 313–366, CSLI Publications.

- Ladd, D. Robert. 1986. Intonational phrasing: the case for recursive prosodic structure. *Phonology Yearbook* 3, 311–340.
- Ladd, D. Robert. 2008. *Intonational Phonology*. Cambridge [a.o.]: Cambridge University Press.
- Ladefoged, Peter. 1971. *Preliminaries to Linguistic Phonetics*. Chicago [a.o.]: University of Chicago Press.
- Lahiri, Aditi. 2000. Phonology: structure, representation, and process. In L. Wheel- don (ed.), *Aspects of Language Production*, Chapter 7, pages 165–226, Psychology Press.
- Lahiri, Aditi, Jongman, A. and Sereno, J. 1990. The pronominal clitic [dər] in Dutch: A theoretical and experimental approach. *Yearbook of Morphology* 3, 115–127.
- Lahiri, Aditi and Plank, Frans. 2008. What linguistic universals can be true of. In S. Scalise, E. Magni and A. Bisetto (eds.), *Universals of Language Today*, volume 76 of *Studies in Natural Language and Linguistic Theory*, pages 31–58, Dordrecht: Springer.
- Lahiri, Aditi and Plank, Frans. 2010. Phonological phrasing in Germanic: the judge- ment of history, confirmed through experiment. *Transactions of the Philological Society* 108(3), 372–398.
- Lahiri, Aditi and Reetz, Henning. 2002. Underspecified recognition. In C. Gussen- hoven, N. Werner and T. Rietveld (eds.), *Labphon 7*, Berlin: Mouton.
- Lahiri, Aditi and Reetz, Henning. 2010. Distinctive features: phonological under- specification in representation and processing. *Journal of Phonetics* 38, 44–59.
- Lapointe, Steven G. 1980. *A Theory of Grammatical Agreement*. Ph.D.thesis, Uni- versity of Massachusetts.
- Leben, William R. 1973. *Suprasegmental Phonology*. Ph.D.thesis, Massachusetts In- stitute of Technology.
- Lee, Hanjung. 2001. Markedness and word order freezing. In P. Sells (ed.), *Formal and Empirical Issues in Optimality Theoretic Syntax*, pages 63–127, CSLI Publi- cations.
- Lehiste, Ilse, Olive, Joseph P. and Streeter, Lynn A. 1976. Role of duration in disam- biguating syntactically ambiguous sentences. *The Journal of the Acoustical Society of America* 60, 1199–1202.
- Levelt, Willem J.M. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

- Levelt, Willem J.M. 1999. Models of word production. *Trends in Cognitive Sciences* 3(6), 223–232.
- Levelt, Willem J.M., Roelofs, Ardi and Meyer, Antje S. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22, 1–75.
- Liang, Percy and Potts, Christopher. 2015. Bringing machine learning and compositional semantics together. *Annual Review of Linguistics* 1(1), 355–376.
- Lieberman, Mark. 1975. *The intonational system of English*. Ph.D.thesis, Massachusetts Institute of Technology.
- Lieberman, Mark and Pierrehumbert, Janet. 1984. Intonational invariance under changes in pitch range and length. In M. Aronoff and R. T. Oehrle (eds.), *Language Sound Structure*, Chapter 10, pages 157–233, MIT Press.
- Lieberman, Mark and Prince, Alan S. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8(2), 249–336.
- Lowe, John. 2011. Ṛgvedic clitics and ‘prosodic movement’. In *Proceedings of LFG11*, Stanford, CA: CSLI Publications.
- Lowe, John J. forthcoming. Clitics: separating syntax and prosody. *Journal of Linguistics* x(y).
- McCarthy, John J. 2007. What is Optimality Theory. *Language and Linguistics Compass* 1(4), 260–291.
- McCarthy, John J. and Prince, Alan S. 1993. Generalized alignment. *Yearbook of Morphology*, pages 79–153.
- McCarthy, John J. and Prince, Alan S. 2001. Prosodic morphology. In A. Spencer and A. M. Zwicky (eds.), *The Handbook of Morphology*, pages 283–305, Oxford [a.o.]: Blackwell.
- McCawley, James D. 1968. *The Phonological Component of a Grammar of Japanese*. The Hague [a.o.]: Mouton.
- McQueen, James M. 2005. Speech perception. In K. Lamberts and R. L. Goldstone (eds.), *Handbook of Cognition*, Chapter 11, pages 255–275, London: SAGE Publications.
- Meinschaefer, Judith. 2005. The prosodic domain of Italian *troncamento* is not the clitic group, arbeitspapier Nr. 118, Linguistics Department, University of Konstanz.

- Mester, Armin and Ito, Junko. 2009. The extended prosodic word. In J. Grijzenhout and B. Kabak (eds.), *Phonological Domains: Universals and Derivations*, pages 135–194, Berlin [a.o.]: Mouton de Gruyter.
- Mohanan, Karuvannur P. 1995. The organization of grammar. In John Goldsmith (ed.), *The Handbook of Phonological Theory*, Chapter 2, pages 24–69, Cambridge, MA: Blackwell.
- Mohanan, Karuvannur Puthanveettil. 1982. *Lexical Phonology*. Ph.D.thesis, Massachusetts Institute of Technology.
- Mohanan, Tara. 1990. *Arguments in Hindi*. Ph.D.thesis, Stanford University.
- Mohri, Mehryar. 2004. Weighted finite-state transducer algorithms: an overview. In C. Martín-Vide, V. Mitrana and G. Păun (eds.), *Formal Languages and Applications*, pages 551–564, Berlin: Springer.
- Moulton, William G. 1947. Juncture in modern standard German. *Language* 23(3), 212–226.
- Mycock, Louise. 2006. *The Typology of Constituent Questions: A Lexical-Functional Grammar Analysis of ‘WH’-Questions*. Ph.D.thesis, University of Manchester.
- Mycock, Louise and Lowe, John. 2013. The Prosodic Marking of Discourse Functions. In *Proceedings of LFG13*, Stanford, CA: CSLI Publications.
- Nespor, Marina and Vogel, Irene. 1986. *Prosodic Phonology*. Dordrecht: Foris.
- Nespor, Marina and Vogel, Irene. 1989. On clashes and lapses. *Phonology* 6(1), 69–116.
- Nordlinger, Rachel. 1998. *Constructive Case: Evidence from Australian Languages*. Dissertations in Linguistics, Stanford, CA: CSLI Publications.
- Nübling, Damaris. 1992. *Klitika im Deutschen: Schriftsprache, Umgangssprache, Alemannische Dialekte*. Tübingen: Gunter Narr Verlag.
- O’Connor, Rob. 2004. *Information Structure in Lexical-Functional Grammar: The Discourse-Prosody Correspondence in English and Serbo-Croatian*. Ph.D.thesis, University of Manchester.
- Odden, David Arnold. 2005. *Introducing Phonology*. New York: Cambridge University Press.
- Pak, Marjorie and Frieser, Michael. 2006. French phrasal phonology in a derivational model of PF. In *Proceedings of NELS 36*, Amherst: University of Massachusetts.

- Parkes, Malcolm B. 1992. *Pause and Effect. An Introduction to the History of Punctuation in the West*. Aldershot: Scolar Press.
- Partee, Barbara H., ter Meulen, Alice and Wall, Robert E. 1990. *Mathematical Methods in Linguistics*, volume 30 of *Studies in Linguistics and Philosophy*. Dordrecht: Kluwer Academic Publishers.
- Pate, David M. 2012. *Second Position Clitics and Subordinate tje Clauses in Pashto*. Masters Thesis, Graduate Institute of Applied Linguistics, Dallas, Texas.
- Penner, Zwi. 1991. Pronominal clitics n Bernese Swiss German and their structural position. Jakob Wackernagel and language acquisition. In H. van Riemsdijk and L. Rizzi (eds.), *Clitics and their Hosts*, pages 253–268, EURO TYP Working Papers.
- Peperkamp, Sharon A. 1997. *Prosodic Words*, volume 34 of *HIL dissertations*. Den Haag: Holland Academic Graphics.
- Pierrehumbert, Janet. 1980. *The phonology and phonetics of English intonation*. Ph. D.thesis, Massachusetts Institute of Technology.
- Pierrehumbert, Janet B. and Beckman, Mary E. 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press.
- Pike, Kenneth L. 1967a. *Language in Relation to a Unified Theory of the Structure of Human Behavior*. The Hague [a.o.]: Mouton, second edition.
- Pike, Kenneth L. 1967b. *Tone languages: A Technique for Determining the Number and Type of Pitch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion*. Ann Arbor: University of Michigan Press, 7. edition.
- Plank, Frans. 2005. The prosodic contribution of clitics: Focus on Latin. *Lingue e Linguaggio* 4(2), 281–292.
- Plank, Frans. 2012. Phrasierungsetüden, unpublished manuscript.
- Pollard, Carl J. and Sag, Ivan A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- Price, Patti, Ostendorf, Mari, Shattuck-Hufnagel, Stefanie and Fong, Cynthia. 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America* 90(6), 2956–2970.
- Prince, Alan and Smolensky, Paul. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford, Malden: Blackwell.

- Pyle, Charles. 1972. On eliminating BM's. In P. Peranteau, J. Levi and G. Phares (eds.), *Papers from the eighth regional meeting of the Chicago Linguistic Society*, pages 516–532, Chicago: Chicago Linguistic Society.
- Radonović-Kocić, Vesna. 1996. The placement of Serbo-Croatian clitics: a prosodic approach. In A. L. Halpern and A. M. Zwicky (eds.), *Approaching Second: Second Position Clitics and Related Phenomena*, Chapter 12, pages 429–445, Stanford, CA: CSLI Publications.
- Roberts, Taylor. 1997. The optimal second position in Pashto. In G. Booij and J. van de Weijer (eds.), *Phonology in Progress - Progress in Phonology: HIL Phonology Papers III*, pages 367–401, Holland Academic Press.
- Roberts, Taylor. 2000. *Clitics and Agreement*. Ph.D.thesis, Massachusetts Institute of Technology.
- Roelofs, Ardi. 1997. The WEAVER model of word-form encoding in speech production. *Cognition* 64(3), 249–284.
- Rose, Sarah. 2014. Degema. In D. Nurse, J. Hewson and S. Rose (eds.), *Verbal Categories in Niger-Congo*, Unpublished. <http://www.mun.ca/linguistics/nico/> (8.1.2014).
- Rotenberg, Joel. 1978. *The Syntax of Phonology*. Ph.D.thesis, Massachusetts Institute of Technology.
- Rubach, Jerzy. 1981. *Cyclic Phonology and Palatalization in Polish and English*. Ph.D.thesis, Wydawnictwa Uniwersytetu Warszawskiego, Warsaw.
- Sadler, Louisa. 1997. English auxiliaries as tense inflections, to appear in *Special Issue of Essex Research Reports produced on the occasion of the retirement of Keith Brown*.
- Sadock, Jerrold M. 1991. *Autolexical Syntax: A Theory of Parallel Grammatical Representations*. Chicago: The University of Chicago Press.
- Salton, Gerard. 1971. *The SMART Retrieval System – Experiments in Automatic Document Processing*. Upper Saddle River, NJ: Prentice Hall.
- Samuels, Bridget D. 2011. *Phonological Architecture: A biolinguistic Perspective*. Oxford [a.o.]: Oxford University Press.
- Scheer, Tobias. 2008. Why the prosodic hierarchy is a diacritic and why the interface must be direct. In J. Hartmann, V. Hegedüs and H. van Riemsdijk (eds.), *Sounds of Silence: Empty Elements in Syntax and Phonology*, pages 145–192, Elsevier.

- Scheer, Tobias. 2011. *A Guide to Morphosyntax-Phonology Interface Theories: How Extra-phonological Information is Treated in Phonology since Trubetzkoy's Grenzsignale*. Berlin: De Gruyter Mouton.
- Scheer, Tobias. 2012. *Direct Interface and One-Channel Translation*. Berlin: De Gruyter Mouton.
- Seidl, Amanda. 2001. *Minimal indirect reference – a theory of the phonology-syntax interface*. New York [a.o.]: Routledge.
- Selkirk, Elisabeth O. 1972. *The phrasal phonology of English and French*. Ph. D. thesis, Massachusetts Institute of Technology.
- Selkirk, Elisabeth O. 1978. On prosodic structure and its relation to syntactic structure. In T. Fretheim (ed.), *Nordic Prosody II*, pages 111–140, Tapir.
- Selkirk, Elisabeth O. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.
- Selkirk, Elisabeth O. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3, 371–405.
- Selkirk, Elisabeth O. 1995. The prosodic structure of function words. In J. N. Beckmann, L. W. Dickey and S. Urbanczyk (eds.), *Papers in Optimality Theory*, University of Massachusetts: Department of Linguistics.
- Selkirk, Elisabeth O. 2009. On clause and intonational phrase in Japanese: the syntactic grounding of prosodic constituent structure. *Gengo Kenkyu* 136, 35–73.
- Selkirk, Elisabeth O. 2011. The syntax-phonology interface. In J. Goldsmith, J. Riggall and A. C. L. Yu (eds.), *The Handbook of Phonological Theory*, pages 435–484, Malden, MA: Blackwell.
- Selkirk, Elisabeth O. and Tateishi, Koichi. 1988. Constraints on minor phrase formation in Japanese. In *Proceedings of the 24th annual meeting of the Chicago Linguistics Society*, pages 316–336, Chicago: Chicago Linguistic Society.
- Sells, Peter. 2001. Introduction. In P. Sells (ed.), *Formal and Empirical Issues in Optimality Theoretic Syntax*, CSLI Publications.
- Shattuck-Hufnagel, Stephanie, Ostendorf, Mari and Ross, Ken. 1994. Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics* 22, 357–388.
- Sievers, Eduard. 1901. *Grundzüge der Phonetik: zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen*. Leipzig: Breitkopf & Härtel, fifth edition.

- Silverman, Kim, Beckman, Mary, Pitrelli, John, Ostendorf, Mari, Wightman, Colin, Price, Patti, Pierrehumbert, Janet and Hirschberg, Julia. 1992. TOBI: A standard for labeling English prosody. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, Banff.
- Spencer, Andrew. 1991. *Morphological Theory – an Introduction to Word Structure in Generative Grammar*. Oxford [a.o.]: Blackwell.
- Spencer, Andrew. 2007. Morphophonological operations. In A. Spencer and A. M. Zwicky (eds.), *The Handbook of Morphology*, pages 123–143, Hoboken, NJ: Wiley-Blackwell.
- Spencer, Andrew and Luís, Ana R. 2012. *Clitics – an Introduction*. Cambridge: Cambridge University Press.
- Sulger, Sebastian, Butt, Miriam, King, Tracy Holloway, Meurer, Paul, Laczkó, Tibor, Rákosi, György, Dione, Cheikh Bamba, Dyvik, Helge, Rosén, Victoria, De Smedt, Koenraad, Patejuk, Agnieszka, Cetinoglu, Ozlem, Arka, I Wayan and Mistica, Meladel. 2013. ParGramBank: The ParGram Parallel Treebank. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 550–560, Sofia, Bulgaria: Association for Computational Linguistics.
- Sweet, Henry. 1876. *Words, Logic and Grammar*. London: Asher.
- Sweet, Henry. 1904. *Elementarbuch des gesprochenen Englisch*. Leipzig: Tauchnitz [u.a.], third edition.
- Tegey, Habibullah. 1977. *The Grammar of Clitics: Evidence from Pashto and other Languages*. Ph.D.thesis, University of Illinois.
- Tegey, Habibullah. 1979. Ergativity in Pushto. In I. Rauch and G.F. Carr (eds.), *Linguistic Method: Essays in Honor of Herbert Penzl*, pages 369–418, The Hague: Mouton.
- Tegey, Habibullah and Robson, Barbara. 1996. *A Reference Grammar of Pashto*. Washington, D.C.: Center for Applied Linguistics.
- Toivonen, Ida. 2003. *Non-Projecting Words: A Case Study of Swedish Verbal Particles*. Dordrecht: Kluwer.
- Trubetzkoy, Nikolai S. 1939/1977. *Grundzüge der Phonologie*. Göttingen: Vandenhoeck & Ruprecht, 6. edition.
- Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30(2), 219–255.

- Truckenbrodt, Hubert. 2007. The syntax-phonology interface. In P. de Lacy (ed.), *The Cambridge Handbook of Phonology*, Chapter 18, pages 435–456, Cambridge: Cambridge University Press.
- UCLA Phonetics Lab Archive. 2007. *Degema – dataset 4, recorded 1994*. Los Angeles, CA: UCLA Department of Linguistics. <http://archive.phonetics.ucla.edu/Language/DEG/deg.html> (10.01.2015).
- van der Hulst, Harry. 2010. A note on recursion in phonology. In H. van der Hulst (ed.), *Recursion and Human Language*, pages 301–242, Berlin: Mouton de Gruyter.
- van der Leeuw, Frank. 1995. Cliticization, stress and phonological words in European Portuguese: an optimal(ity) approach. *Probus* 7(1), 31–68.
- Vennemann, Theo. 1988. *Preference Laws for Syllable Structure and the Explanation of Sound Change*. Berlin [a.o.]: Mouton de Gruyter.
- Vogel, Irene and Kenesei, István. 1990. Syntax and semantics in phonology. In Sharon Inkelas and Draga Zec (eds.), *The phonology-syntax connection*, pages 339–363, Chicago [a.o.]: University of Chicago Press.
- Wackernagel, Jacob. 1892. Über ein Gesetz der indogermanischen Wortstellung. *Indogermanische Forschungen* 1, 333–436.
- Wagner, Michael. 2005. *Prosody and Recursion*. Ph.D.thesis, Massachusetts Institute of Technology.
- Wescoat, Michael T. 2002. *On Lexical Sharing*. Ph.D.thesis, Stanford University.
- Wescoat, Michael T. 2005. English nonsyllabic auxiliary contractions: an analysis in LFG with Lexical Sharing. In *Proceedings of LFG05*, Stanford, CA: CSLI Publications.
- Wescoat, Michael T. 2009. Udi person markers and lexical integrity. In *Proceedings of LFG09*, Stanford, CA: CSLI Publications.
- Wheeldon, Linda. 2000. Generating prosodic structure. In L. Wheeldon (ed.), *Aspects of Language Production*, pages 249–274, Psychology Press.
- Wheeldon, Linda and Lahiri, Aditi. 1997. Prosodic units in speech production. *Journal of Memory and Language* 37, 356–381.
- Young-Scholten, Martha. 1993. *The Acquisition of Prosodic Structure*, volume 304 of *Linguistische Arbeiten*. Tübingen: Niemeyer.

-
- Zec, Draga. 2007. The syllable. In P. de Lacy (ed.), *The Cambridge Handbook of Phonology*, Chapter 8, pages 161–194, Cambridge [a.o.]: Cambridge University Press.
- Zec, Draga and Inkelas, Sharon. 1990. Prosodically constrained syntax. In Sharon Inkelas and Draga Zec (eds.), *The Phonology-Syntax Connection*, pages 365–378, Chicago: University of Chicago Press.
- Zwicky, Arnold and Pullum, Geoffrey K. 1983. Clitization vs. inflection: English *n't*. *Language* 59(3), 502–513.
- Zwicky, Arnold M. 1977. *On Clitics*. Bloomington: Indiana University Linguistics Club.