Marika Butskhrikidze

The Consonant Phonotactics of Georgian

The central topic of this thesis is the study of Georgian consonant sequences, e.g. forms of the CCC type. It demonstrates that the complexity of the Georgian consonant clusters is related to morphological complexity and to processes of vowel reduction and complex segment formation. Thus, the Georgian 'complex' CCC sequences are derived from structures of the CVCVCV type.

For the representation of the consonant phonotactics of Georgian, a phonological hierarchy is introduced in which the stem domain is intermediate between the segment and the word domain. It is proposed that a number of phonological principles, the Obligatory Contour Principle, the Sonority Sequencing Principle, the Syllable Contact Law and the Principle of Resolvability are instantiations of a single principle, the Balancing Principle. They govern the consonant co-occurrence restrictions and function at the stem or the word domain, depending on the language type.

The hypothesis that a language which has C_iC_j clusters will also have C_iVC_j stems is proposed and verified on the basis of Georgian data. This relates to the claim that Georgian clusters are maximally biconsonantal.

The Gradual Consonant Analysis, based on different types of evidence - (i) paradigmatic and syntagmatic, (ii) historical, (iii) phonetic and (iv) comparative - provides a direct and clear link between empirical structures and theoretical constructs, and explains why consonants form complex structures in Georgian.

This book is of interest to linguists studying phonotactics, phonetics, the phonology-morphology interface and Georgian.

: LOT
Netherlands
Graduate
School of
Linguistics

ISBN 90-76864-24-1

: LOT 63
Marika Butskhrikidze The Consonant Phonotactics of Georgian

Marika Butskhrikidze

The Consonant Phonotactics of Georgian

Universiteit Leiden
Centre for Linguistics

HIL

: LOT
Netherlands
Graduate
School of
Linguistics

Landelijke Onderzoekschool Taalwetenschap

The Consonant Phonotactics of Georgian

Published by
LOT
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6006
fax: +31 30 253 6000
e-mail: lot@let.uu.nl
http://www.let.uu.nl/lot/

Cover illustration: The oldest specimen of Georgian writing, found in the Bolnisi Sioni Temple (5th century).

ISBN 90-76864-24-1
NUGI 941

The Consonant Phonotactics of Georgian

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus Dr. D.D. Breimer,
hoogleraar in de faculteit der Wiskunde en
Natuurwetenschappen en die der Geneeskunde,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 6 november 2002
te klokke 16.15 uur

door

Marika Butskhrikidze

geboren te Tbilisi (Georgië)
in 1971

Promotiecommissie

promotores:	Prof. dr. C.J. Ewen
	Prof. dr. H.G. van der Hulst (University of Connecticut)
co-promotor:	Dr. J.M. van de Weijer
referent:	Prof. dr. E.V. Hume (Ohio State University)
overige leden:	Prof. dr. V.J. van Heuven
	Dr. H.J. Smeets

ვუძღვნი ჩემს მშობლებს
[To my parents]

Contents

Acknowledgements

It is disappointing to be unable to thank the people who have contributed the most to the completion of this book. Therefore, I would like to express my gratitude collectively to all members of the ULCL for creating a friendly and stimulating environment in which both "colourless green ideas" and individualism are appreciated.

I am very grateful to a number of people who have read the whole thesis or separate chapters and provided me with helpful comments and suggestions: Crit Cremers, Rob Goedemans, Jan Kooij, Anthi Revithiadou, Nancy Ritter and Grażyna Rowicka. I am also grateful to people with whom I have discussed issues considered in this thesis: Rusudan Asatiani, Laura Benua, Geert Booij, Bruce Hayes, Larry Hyman, Marika Jikia, Paul Kiparsky, James McQueen, Irine Melikishvili, Jos Pacilly, Sang Jik Rhee, Eter Soselia and Miklós Törkenczy.

My conference trips and two sets of fieldwork in Georgia would not have been possible without the financial support of the Holland Institute of Generative Linguistics (HIL), the Netherlands Graduate School of Linguistics (LOT), the Leiden University Fund (LUF), the Dutch Organisation for Scientific Research (NWO) and the University of Leiden Centre for Linguistics (ULCL). Their contributions are gratefully acknowledged. I am also thankful to NWO for a language correction grant and to Mark de Vos for thoroughly checking the entire manuscript. José Birker, our secretary, has also helped me in various ways and I am very grateful to her.

Thanks are due to my friends, Bert Botma, Nancy Chongo Kula, Claartje Levelt, Sang Jik Rhee, Grażyna Rowicka, Erik Jan van der Torre and Marina Tzakosta for making phonological life interesting and lively at the ULCL.

I would also like to express my gratitude to people who contributed to my development as a linguist.

I am very fortunate to have received guidance from my student years onwards from such an accomplished academic as Thomas Gamkrelidze. Without his constant support and care I would not have been able to continue research in linguistics and I am very grateful that he has shown his confidence in me.

It was a great experience to be a part of the Department of Structural and Applied Linguistics at the Tbilisi State University and the Department of Language Typology at the Institute of Oriental Studies. Many thanks are due to all my teachers and colleagues in both places. I owe special thanks to Zurab Sardschweladse and Bakar Gigineishvili for giving me an opportunity to work with a great team of Kartvelologists at the Laboratory of Kartvelian languages at the Sulkhan-Saba Pedagogical Institute.

I was lucky to meet such an exceptional teacher and personality as Miklós Törkenczy, my tutor in Budapest. He was the first who introduced me to generative phonology and thanks to his persistence, patience and kindness I learned to look at things differently.

Sometimes, the right words said at the right time can change a life radically, and I am very thankful to my friends Jan Don, Jonathan Ginzburg and Enric Vallduví for their support and encouragement to pursue my studies abroad.

My linguistic life has been connected with my best friend, Nino Amiridze, ever since the very first day we started doing linguistics; I can never thank her enough for all the inspirations and support. Nin, good luck with your beautiful family and with your book!

In my family love has no boundaries and it feels much stronger even now when I am abroad. I would like to thank my mother, Guliko, my aunt, Eteri, my sister, Nino, and my brother, Dima, for their faith in me and for the beautiful sense of bond that we have. I would also like to thank my sister-in-law, Merita, for being there unconditionally for me.

And finally, any formal thanks to Bedri and Teuta would be inappropriate; they make my home ground far more true and fascinating than any linguistic ground could ever be.

Abbreviations and Symbols

ABL	ablative
ADJ	adjective
ACC	accusative
aff.	affix
ASP	aspect
BP	Balancing Principle
C	consonant
CAUS	causative
COMP	Complementiser
CP	Compensatory Principle
DAT	dative
ERG	ergative
Eng.	English
F	foot
Fr.	French
GCA	Gradual Consonant Analysis
Georg.	Georgian
GEN	genitive
Ger.	German
gram.	grammatical
IMP	imperative
INF	infinitive
INST	instrumental
HSA	Headless Syllable Analysis
lex.	lexical
LOC	locative
morph.	morpheme
N	noun
N.AG.	nomen agentis
NOM	nominative
OCP	Obligatory Contour Principle
PERF	perfective
PL	Prosodic Licensing
PL	plural
POSS	possessive
PR	Principle of Resolvability

PREF	prefix
PRES	present tense
PREV	preverb
PRT	participial suffix
PrWd	prosodic word
Russ.	Russian
σ	syllable
SCA	Syllabified Consonant Analysis
SCL	Syllable Contact Law
SG	singular
Sp.	Spanish
SSP	Sonority Sequencing Principle
SUFF	suffix
syl.	syllable
THEM	thematic
V	vowel
V	verb
VOC	vocative
1^{st}	first person
2^{nd}	second person
3^{rd}	third person
ˊ	primary word accent
ˋ	secondary word accent
ː	length
#	boundary marker
.	syllable boundary
-	morphological boundary
*	unattested form
☞	winning candidate
/–/	phonological form
[–]	phonetic form

1 Preliminaries

The whale has a two-fold distinction among the fishes:
first, when seen from a distance,
it looms large among them
and, secondly, on close examination
it is found to be no fish at all.
M.F. Meyer[1]

1.0. Introduction: the 'problem'

"… *science starts only with problems*. Problems crop up especially when we are disappointed in our expectations, or when our theories involve us in difficulties, in contradictions" (Popper 1963:222).

The consonant sequences of Georgian, such as /prckvn/, /mc'vrtn/, /brt'χ'/ in words like /prckvna/ 'to peel', /mc'vrtneli/ 'trainer', /brt'χ'eli/ 'flat', have long been an object of interest to professional linguists as well as interested laymen. The interest is mainly due to two characteristics of the Georgian consonant sequences: (i) their unusual length – the number of consonants in word-initial position of underived words can be up to six – and (ii) their unusual constituency – at first sight consonants in sequences do not appear to show any organisational principle at work. Both length and constituency might lead us to the assumption that consonant sequences, and consequently the phonotactic organisation of the whole word, are arbitrary in Georgian. Arbitrary phonological structures present a challenge for linguists who consider language as a system, i.e. a structured and organised entity. Thus Georgian consonant sequences are problematic for linguistic theories assuming a systemic nature of language.

One can think of two ways to resolve the 'problem': (i) to say that the complexity is just an artefact of orthography or (ii) to attribute such complex structures to some phenomenal endowment shared by speakers of Georgian. As far as the former is concerned, native speakers have no difficulty in pronouncing words like the above, with all the consonants. As for the latter, equally complex patterns of consonant clustering in, for instance, Polish, Berber and Bella Coola mean that this phenomenon has a broader scope and consequently place it among the problematic objects of study in a general theory of language and language modelling.

[1] The citation is taken from the book by Fodor & Katz (1964).

Previous analyses of Georgian consonant clusters, for instance, Nepveu (1994) and Toft (1999), deny the arbitrariness of consonant clustering (see Chapter 7). Both analyses are couched within the Generative framework. This thesis also argues that consonant sequences in Georgian are not random. However, it offers an analysis which is based on different theoretical and methodological premises. They are introduced below. Although I consider ontological, epistemological and methodological premises to be interrelated and interdependent, I present them in separate sections, for ease of exposition.

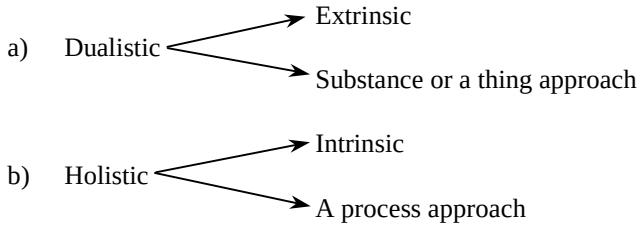
1.1. Ontological premises

The long history of linguistic thought can, in my view, be recapitulated as the study of dualistic bifurcations: *appearance* vs. *essence*, *abstract* vs. *concrete*, *whole* vs. *part*, *invariant* vs. *variant*, *regular* vs. *irregular*, *form* vs. *meaning*, *unmarked* vs. *marked*, *lexicon* vs. *grammar*, *specific* vs. *universal*, etc. In order to achieve the two main goals of linguistic science, classification and explanation, a researcher has to address these linguistic opposites and unequivocally determine their place and function in the language faculty. This is not an easy task, taking into consideration that borderlines between the alternates are often blurred.

There are at least two stands which a researcher can take with respect to the way the members of the oppositions are related: (i) dualistic and (ii) holistic.[2] The former implies that the members of the oppositions are different in nature and represent two modes of existence, i.e. are extrinsic; the latter implies that although the members of the opposition are different, they are interdependent, i.e. are intrinsic. The Generative framework, based on Cartesian philosophy, takes the dualistic standpoint by regarding alternates as different planes of investigation. Consequently, Generative theories make a sharp distinction between form and meaning, lexicon and grammar, etc. (see Contini-Morava & Tobin 1984, Stubbs 1996 and Schönefeld 1997 for discussion). I take the holistic position, because I consider language to be a social phenomenon. In an epistemological context, this means that the subject, which possesses a language, and the object, the reality, are seen as a unity, and consequently the knowledge of language is a relation, since language is a part of reality. The opposition *dualistic* vs. *holistic* can be characterised as follows:

[2] It seems superfluous for the discussion here to consider a monistic position, which assumes that all of reality is of one kind.

(1)



(Israel 1979:8)

The difference depicted in (1) has serious consequences for the way language and linguistic opposites are defined. The holistic approach adopted in this thesis suggests that language is not a static, closed, fixed system, but instead an open, dynamic one. With respect to linguistic opposites this implies that opposites, for example marked vs. unmarked, lexicon vs. grammar, are not fixed, but are instead variables, which are highly dependent on the context (see e.g. Melikishvili 1976, who argues for context-dependent markedness in phonology, and Lehmann 1986 and Langacker 1988, who advocate the conventional view of a division between lexicon and grammar in language). In the following sections, three linguistic oppositions, *marked* vs. *unmarked*, *form* vs. *meaning* and *lexicon* vs. *grammar*, are discussed to illustrate that the alternates of the oppositions are relational and cannot be viewed independently from each other.

1.1.1. Marked vs. unmarked

Consider an example from the study of word-level phonotactics, viewed in terms of markedness. The experiment discussed here is a 'thought experiment' described by Brasington (1997). Imagine two languages (call them language A and language B), which differ in their permissible phoneme sequences (i.e. phonotactics) but not in their phonemic inventories (for instance, both languages have only the consonants /p t k m s/ and the vowels /a i u/). Language A requires words to be made up of alternating sequences of consonants and vowels. Furthermore, words must begin with a consonant and end with a vowel. As a result of this constraint, the vocabulary of language A contains words such as /palaka/, /masitu/, etc. The words in such a language would tend to be rather 'long', for only in this way is it possible to provide for a sufficient variety of different word-forms. Language B, by contrast, allows words with groups of consonants at the beginning and the end. Obeying these regularities, language B is able to distinguish many more 'short' single syllable words than language A. Language B can take advantage of quite complex single syllable forms like /skumps/, /stukt/ to stock up its lexicon. It would be difficult to find a language which would exactly match language A or language B, but one could say that English, for example, tends towards a language B type and Rennellese, a language of the Solomon Islands, tends towards a language A type. The patterns of these two types of languages are exaggerated in this example. One could also imagine languages which have both types of structural patterns. Georgian is one such language. In a statistical account of the Georgian consonant clusters, Bush (1997) finds an inter-

esting generalisation: "the fewer the syllables in a word, the longer the consonant clusters tend to be; the more syllables, the shorter the clusters ... this generalization seems to reflect an interesting principle at work: there is a trade-off between the number of syllables and the length of consonant clusters" (Bush 1997:49).

In the two extreme cases of language A and language B, the following two questions arise:

(2) a) Is the phenomenon found in language A different from the one in language B?
b) How can the difference be characterised in terms of markedness, assuming that complex structures are assigned marked status?

My answer to the question in (2a) is that both languages are characterised by a single phenomenon. The underlying principle governing the phonotactic organisation of both languages is that of communicative adequacy, which in turn relies on having a sufficiently 'stocked' lexicon.

In order to answer the question in (2b), it is important to notice that language A is complex in terms of number of syllables, while language B is complex in terms of consonant groupings. Thus, being marked or unmarked depends on the terms the opposition is being defined in. If number of syllables is taken as the criterion, then language A is more marked, while if consonant grouping is considered as the criterion, language B appears to be more marked. Thus, the nature of the opposition marked vs. unmarked is relational, or context-dependent (see (14)).

(3) *Phonotactic markedness in terms of two variables*

Variables Language types	Number of syllables	Consonant groupings
Language A (CVCVCV)	Marked	Unmarked
Language B (CCCV)	Unmarked	Marked

Languages have different appearances, which are reflected in their directly observable surface forms, but when closely examined the phonological characteristics of these forms are universal (perhaps because they aspire to one goal, as part of a communicatively adequate system). In Modern Georgian, 'complexity' of the type attested in language B appears to predominate. This thesis demonstrates that the complexity of the Georgian consonant clusters is the result of processes of vowel reduction and complex cluster formation. Thus, the Georgian 'complex' CCC sequences are in fact transformations of structures of the CVCVCV type.[3] Such trans-

[3] Within the Strict CV framework of Government Phonology Lowenstamm (1996) argues that all languages only have CV 'syllables' or units, and structures like CCV, e.g. in English or in Dutch, should be analysed as CVCV (with an empty nucleus in between).

formations are expected if language is seen not as static, fixed,[4] closed system, but as a dynamic, open system (see section 1.4 for further discussion).

1.1.2. Form vs. meaning

In my opinion, the opposition crucial to the study of surface consonant clustering and phonotactics is the opposition between lexicon and grammar. The discrepancy is most prominently demonstrated in the studies of form-meaning mapping. The studies are summarised briefly below.

> ... language is a system of signs, and linguistics is part of the science of signs, or SEMIOTIC (Saussure's *sémiologie*). The ancient definition of the sign – "aliquid stat pro aliquo"[5] – has been resurrected and proposed as still valid and productive. Thus the essential property of any sign in general, and of any linguistic sign in particular, is its twofold character: every linguistic unit is bipartite and involves two aspects – one sensible and the other intelligible – or, in other words, both a *signans* (Saussure's *signifiant*) and a *signatum* (*signifié*). These two constituents of any linguistic sign (and of any sign in general) necessarily presuppose and require each other (Jakobson 1971:103).

The issue of form-meaning mapping is vast and represents the philosophical backbone of any linguistic theory. I will not attempt a detailed summary here but simply mention some approaches that are relevant to my understanding of the relationship between form and meaning, and its implications for the study of word-level phonotactics.

I discuss previous studies concerning the nature of the relation between form and meaning, such as de Saussure (1916), Pierce (1931), Jakobson (1971) and Gamkrelidze (1974). The period between de Saussure (1916) and Gamkrelidze (1974) represents a move from a segregationalist[6] to an integrationalist approach. The integrationalist approach is the one advocated in the thesis.

In his general classification of signs, Pierce (1931) distinguishes between *signans* and *signatum*. The differences seen in the relationship between *signans* and *signatum* provide a basis for distinguishing three types of signs:

---

[4] Language is seen as a fixed, closed system (just like mathematical systems) within the generative framework. See e.g. Chomsky (1968) who says that it is reasonable to suppose that a generative grammar is a system of many hundreds of rules of several different types, organised in accordance with certain fixed principles of ordering and applicability and containing a certain fixed substructure which, along with the general principles of organisation, is common to all languages.

[5] Something that stands to imply the existence of something else.

[6] Segregationalism is the integrational linguistic cover-term for mainstream linguistic thinking which assumes a clear and permanent separation between what is linguistic and what is extralinguistic. It also assumes robust distinctions between what is text and what is context, between language and parole, performance and competence, etc. Integrational linguistics questions all these simplificatory oppositions (Toolan 1999).

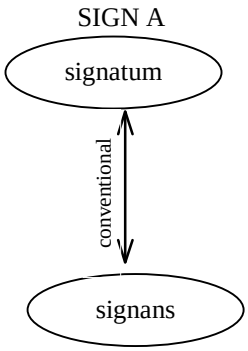
(4) a) *Icons*, i.e. signs in which the signans is characterised by a certain degree of factual similarity with the signatum, being its formal reflection or repetition, e.g. the picture of an animal as an iconic sign of the animal itself.[7]

b) *Indices*, i.e. signs in which the signans is attached to the signatum by a causal or any other relationship, e.g. smoke as a sign-index of fire.

c) *Symbols*, i.e. signs in which the signans is linked with the signatum by convention, where this connection does not depend on the presence or absence of any resemblance or physical contiguity, being arbitrary and realised because of a certain rule of convention, e.g. linguistic signs (Pierce 1931).

De Saussure's (1916) linguistic signs consist of a signifier and signified. These do not occur separately. De Saussure used the analogy of a piece of paper – the sign is like the sheet of paper with the signifier on one side and the signified on the other. De Saussure said that the signifier is 'the form which expresses the word' and the signified is the meaning. Signifier and signified are arbitrarily linked, or as one might say, the link between form and meaning is conventional. The proposal is illustrated in (5).

(5) *Form-meaning mapping in the linguistic sign (De Saussure 1916)*



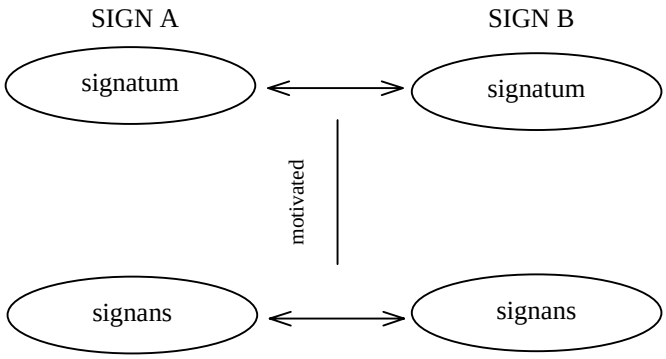
Jakobson (1971) showed that Pierce's trichotomy of signs – icon, index and symbol – may be reformulated as the junction of two dichotomies: similarity/contiguity and factual (existential)/imputed.[8] Jakobson extensively discusses very important aspects

[7] Iconicity is present in the phenomenon of reduplication discussed in Chapter 5. According to Jakobson, in reduplicative forms there is the iconic relation between the linear word order and the meaning which it denotes (e.g. plurality, iteration, duration).

[8] Jakobson redefined 'conventional' as 'imputed'. In his analysis, the icon evidences a factual similarity relation between signans and signatum. The index evidences a factual contiguity relation, and the symbol an imputed contiguity relation; in addition, there is the 'artifice' (not included in Pierce's schema), an imputed similarity relation. The artifice is particularly evident in such non-linguistic semiotic systems as music, but it is also especially important in poetry, where equivalence based on similarity becomes the

of similarity and contiguity relations for the relations between signs with respect to signans and signatum. The so-called 'horizontal' relations between signs show that the internal structure of the signans of a given sign may be motivated from the point of view of the other signs in the same system. "Thus, diverse relations between the signata on the plane of content (e.g. the relation of quantity, 'one' vs. 'many'; of proximity in space and time, 'near' vs. 'far', 'early' vs. 'late'; of size, 'large' vs. 'small'; the relation of similarity/dissimilarity; the relations of contiguity, kinship etc.) are expressed in the corresponding signantia by specific correlations of phonemic similarity, by phonemic alternations (juxtapositions), by phonemic length of the interrelated words, and by other universal properties of the plane of expression of language" (Gamkrelidze 1974:105–106). Thus, the phonetic similarity of such paired numerals as Russ. *devjat'* 'nine' ~ *desjat'* 'ten', Ger. *zwei* 'two' ~ *drei* 'three', which arose as a result of an analogy of one form to the other, is caused by the existence of a certain relationship at the level of the signata: the relation of immediate neighbourhood between numerals is reflected in a partial phonetic resemblance between the corresponding signantia. The same kind of relationship can be illustrated by way of kinship terms, e.g. Eng. *father* ~ *mother* ~ *brother* (Gamkrelidze 1974). Jakobson's (1971) proposal is illustrated in (6).

(6) *Form-meaning mapping in the linguistic sign (Jakobson 1971)*



Gamkrelidze (1974) proposes that:

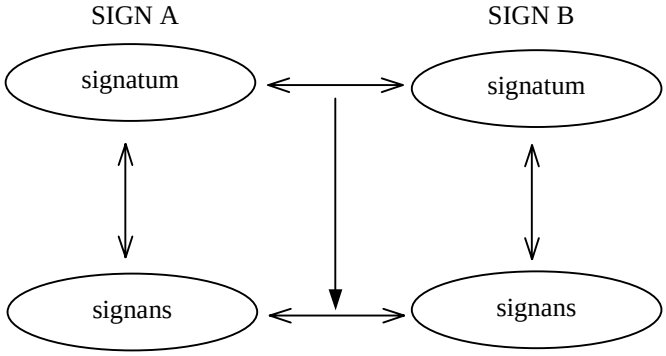
> The nature of the verbal sign, just like the sign in any semiotic system, must be specified through the relationship between the signans and signatum, as well as through the relations of the given sign to the other signs of the system, both at the level of the signata and at that of the corresponding signantia. In defining a sign of a semiotic system, then, we must consider not only the 'vertical' relationship between the two components of a sign taken in isolation, but also the

constitutive device of the sequence, and where an equivalence of signantia between two rhyming words may entail an imputed similarity relation (=artifice) for the corresponding signata (Jakobson 1971, see also Waugh 1978).

> twofold 'horizontal' relations existing between the respective components of the interrelated signs. The Saussurean thesis of the arbitrariness of the sign is partial and incomplete, in that it specifies only the 'vertical' relations, disregarding the nature of the 'horizontal' ones. On the other hand, the opposite views on the motivated, iconic nature of the bond between the signans and signatum, as maintained by the adversaries of the Saussurean thesis, involve exclusively the sphere of the 'horizontal' relations and cannot refer to the 'vertical' relationship, which is characterised in principle by arbitrariness and conventionality (Gamkrelidze 1974:102).

Gamkrelidze proposes that if the verbal sign is conceived of as a unity of the relations between the 'vertical' and 'horizontal' relations, then the relations between signans and signatum present themselves not as contradictory, but as complementary (in Bohr's 1948, 1958 sense of the term), specifying with the necessary completeness the essence of the verbal sign. The idea of complementariness is schematised in (7).

(7) *Form-meaning mapping in the linguistic sign (Gamkrelidze 1974)*



To summarise, the nature of the form-meaning mapping of the linguistic sign seems to be not entirely arbitrary after all (Gamkrelidze 1974). Characteristics of the mapping depend on where we study the correlation, in the grammar or in the lexicon. As shown above, arbitrariness is not maintained in the realm of grammar (Jakobson 1971), while the mapping tends to be arbitrary in the realm of the lexicon (de Saussure 1916). Crucial to this thesis is Gamkrelidze's claim about the complementariness of the horizontal and vertical form-meaning mappings in the linguistic sign. The result is that form and meaning are not seen as completely independent domains. The claim has implications which are for the study of phonotactics. Unlike previous approaches, in which the formal side of a language, i.e. phonotactics, is interpreted in terms of formal units, i.e. the syllable, the foot, the onset and the nucleus (see Chapter 7), I argue that it is the meaning-bearing units, i.e. morphological units, which account for the phonotactic patterns of a language. Such an approach to

the study of phonotactics bridges the gap between phonology and other modules of a language, e.g. morphology. I will demonstrate in this thesis that the study of phonotactics gains explanatory power when morphology, for instance, is taken into account, and leaves formal structures unintelligible when it is ignored.

In line with this reasoning, I argue in this thesis that the lexical part of a word, the stem, constitutes the domain of phonotactic generalisations, and in fact most of the patterns of consonant sequences are stated with respect to the stem domain (see Chapter 2).

1.1.3. Lexicon vs. grammar

Just as the adequate characterisation of the linguistic sign is not a completely resolved issue, neither is language, with its conventional division into lexicon and grammar. This short excursion into the study of the form-meaning mapping has shown the importance of the division between lexicon and grammar. The thesis shows that in order to study the surface phonotactics of a language, it is crucial to establish exactly where the dividing line is between lexicon and grammar (for a detailed discussion see Chapter 2). Both the phonotactics and the dividing line between lexicon and grammar are language-specific. Nevertheless, there are general characteristics associated exclusively with the grammar and the lexicon. Consider the following two cross-linguistic generalisations.

Firstly, it is a commonly acknowledged cross-linguistic observation that non-lexical morphemes, i.e. affixes, obey specific phonotactic constraints that do not hold for lexical morphemes (Jakobson 1949, Nida 1949, Hyman 1975, Booij 1985 and Zubkova 1988, among others). For instance, native prefixes in Dutch consist of at most one syllable, and native suffixes of at most two, of which only one may contain a full vowel, whereas lexical morphemes do not conform to these restrictions (Booij 1977:22–23). Of the twenty-three consonants in spoken Czech, only eight are used in inflectional suffixes (Jakobson 1949:108), and of the twenty-eight consonants in Modern Georgian only eight.

Secondly, there is some evidence suggesting that vowels carry grammatical meaning (information), while consonants carry lexical meaning. “The traditional etymology of the Sanskrit name for consonant, *vyañjana*, as ‘revelative’ seems to carry the suggestion that the consonants rather than the vowels are responsible for the differentiation of meaning” (Allen 1953:81), and, in fact, the higher informational load of consonants is a phenomenon which finds expression in those alphabetic systems limited to symbols for consonants only. In children’s language, the sense-determinative role of consonants generally antedates that of vowels (i.e. the oppositions within the consonant system appear before those in the vowel system) (Jakobson & Waugh 1979). One finds few languages with obstruent clusters as constituents of a grammatical morpheme. If a language has consonant clusters, they are more likely to appear within a stem, rather than within a grammatical morpheme. This correlation between the stem and obstruent clusters is attested in fixed accent languages (Butskhrikidze 1998a). Butskhrikidze (1998a) proposes the following two generalisations: (i) in languages where word = stem, obstruent clusters commonly

occur in word-final position, while (ii) in languages where word = stem + suffix, obstruent clusters occur in word-initial, not in word-final position. The former generalisation is attested in Armenian, French, Itelmen and Nanay. The latter generalisation is attested in Georgian and Sanskrit. I consider these languages in turn.

In Old Armenian (Grabar) and Modern Armenian the word coincides with the stem and the fixed accent is on the final syllable. Obstruent clusters are permitted only in word-final position, never word-initially. The following examples illustrate word-final obstruent clusters in Grabar (Tumanjan 1971).

(8)	bríšk	'doctor'
	ást	'fat'
	yást	'sacrifice'

In French the word coincides with the stem, the fixed word accent is on the final syllable and obstruent clusters appear only in final position:

(9)	robúst	'strong'
	tríste	'sad'

In Itelmen the word coincides with the stem and the fixed accent falls on the final syllable. Generally, all Itelmen words end in consonants. Obstruent clusters are characteristic of word-final position. There are several biconsonantal combinations found word-initially, but one of the members is always a sonorant (Stebnickij 1934).

In Nanay the word coincides with the stem and the fixed accent is word-final. Obstruent clusters occur only in the final position of a word, never word-initially. The same holds true in all Tungus-Manjuric languages (Avronin 1959).

Let us now examine languages having the word-form stem + suffix.

In Georgian the word consists of a stem + suffix.[9] In nouns the suffix is the nominative case marker *-i*, for example *kac-i* 'man' (*kac-* is the stem and *-i* is the nominative case marker), *c'ign-i* 'book' (*c'ign-* is the stem and *-i* is the nominative case marker). In Georgian the accent is on the first syllable of the word and obstruent clusters occur word-initially, never word-finally.

(10)	xšír-i	'frequent'
	sc'ór-i	'right'

The same can be said of another Kartvelian language, Megrelian.

In Sanskrit words are never represented as bare stems. Accent placement is restricted to the second and third syllables from the end, depending on the weight of the penultimate syllable. Obstruent clusters occur word-initially, never word-finally.

[9] Vowel-final stems are exceptions to this generalisation.

(11) kʃám-a ‘patience’
sthír-a ‘eternal’

Thus, in languages with the word-form stem + suffix obstruent clusters appear word-initially rather than word-finally.

Both generalisations suggest that distribution of obstruent clusters is sensitive to the stem. Generally the two cooccur. The generalisations are summarised in (12).

(12) *The distribution of the stem and obstruent clusters in fixed accent languages*

Position in a word Word structure	Word-initial	Word-final
Word = stem + suffix	clusters attested	clusters not attested
Word = stem	clusters not attested	clusters attested

If the correlation between the stem and obstruent clusters are correct, one would expect obstruent clusters to occur in word-initial position in languages where word = stem, but, as we have seen, some of the languages (e.g. French, Armenian and Itelmen) only allow clusters of which one member is a sonorant, but not obstruent clusters. Here a quite general cross-linguistic observation seems be at work, namely the tendency of allowing complexity at only one of the edges of a word, not both.

The correlation between the stem and consonant clusters is incorporated in the hypothesis proposed in Chapter 2. The hypothesis says that the patterns of consonant clusters are reminiscent of patterns of stems. The hypothesis is tested on Georgian data (see Chapter 3).

Both observations lead to the conclusion that the sensitivity of surface word-level phonotactic patterns (especially those involving consonant clustering) to morphological constituency should be reflected in the analysis of clusters. The issue is intensively discussed in Chapter 2, where the stem is introduced as a phonological domain, occupying a place between the segment and the word in the phonological hierarchy.

1.2. Epistemological premises

One central issue in epistemology is the subject-object relation, i.e. the relation between role of a subject who possesses or produces knowledge and what the knowledge is about, its object. A dualistic approach with a sharp distinction between subject and object views language as something which is about reality, but is not understood to be part of reality itself. A holistic approach assumes that knowledge is a relation between a subject and object in which the subject produces cognitive objects. However, these cognitive products are closely related to a world of objects which exists independently of the subject producing them (Israel 1979:18). Otherwise, knowledge about language is knowledge about reality and especially about social reality. In the former approach, language is a closed system and linguistics is

an autonomous discipline, while in the latter approach language is seen as an open system and linguistics as a non-autonomous discipline (see Derwing 1980 and Harris 1987 for discussion). Furthermore, these two approaches differ in their answers to the genetic questions: how do we obtain knowledge? How did we learn to use language? Two viewpoints are distinguished: mentalistic and non-mentalistic. Proponents of the mentalistic view maintain that linguistic theories aim to explain and predict a speaker's 'tacit' knowledge of a language (Chomsky 1965). Non-mentalistic view argues that theories of language are characterised as being about language and not about mental states or mechanisms underlying the speech behaviour of individual speakers (see Botha 1971 and Ringen 1980 for discussion). Many non-mentalists explicitly construe language as a cultural or social phenomenon. The non-mentalistic viewpoint is the one adopted in this thesis.

1.3. Language as an open system

Considering language as a system implies the recognition of regular interrelations and dependencies, not only between the linguistic units of one and the same module of a language (homogeneous units, e.g. phonological, morphological or syntactic), but between linguistic units from different linguistic modules (heterogeneous units, e.g. those concerning the phonology-morphology or phonology-syntax interfaces). A relevant consequence of this approach for this thesis is to show that in order to account for word-level phonotactics, not only the purely phonological constituency of a word should be taken into account, but also its morphological constituency. The importance of the interaction between the two modules advocated in the thesis is related to the way in which I see the form-meaning nexus, discussed below.

Sapir was one of the first to emphasise that a tendency to isolate phonology and grammar as mutually irrelevant linguistic provinces is unfortunate. There are likely to be fundamental relations between them and their respective histories that we do not yet fully grasp (Sapir 1921). Such an interface can be demonstrated on the basis of diachronic data from Persian (Butskhrikidze 1994, 1996, 1998a). For instance, Butskhrikidze (1996) proposes that there is a correlation between the position of the stem and the placement of fixed accent in a word, assuming that prosodic prominence should be on the part of the word which has the greatest informational load, the stem. Two hypotheses are formulated:

(13) a) Hypothesis: If a language has the basic word-form stem + suffix, the accent will fall on a non-final syllable.
b) Hypothesis: If a language has the basic word-form that coincides with the stem, the accent will fall on the final syllable.

Both hypotheses have been verified for fixed accent languages. The hypothesis formulated in (13a) holds for the languages in (14a), while the hypothesis in (13b) holds for the languages in (14b) (Butskhrikidze 1996, 1998a).

(14) a) Kartvelian (Georgian, Megrelian, Laz); Germanic (Yiddish, German, Icelandic); Jeniseyan (Ket); Finno-Ugric (Estonian, Finish, Karelian, Livonian, Saam); Slavic (Polish); Indic (Sanskrit); Iranian (Old Persian).

b) Iranian (Persian, Beludch, Tadzhic, Kurdish, Jazghulam, Ishkashim); Lezgian (Lezg); Manchu-Tungus (Nanay, Itelmen, Oroch); Armenian (Grabar, Modern Armenian); Turkic (Turkish, Karaim, Turkmen, Uzbek, Gagauz, Kumukh, Tuva, Khakas); Finno-Ugric (Udmurt); Romance (French).

The summary of these implicational generalisations is given in (15).

(15) *The placement of the stem and the accent in fixed accent languages*

The placement of accent Word structure	Final	Non-final
Word = stem + suffix	no	yes
Word = stem	yes	no

It is interesting to test the predictive power of these typological generalisations in language change. The developmental patterns of Persian show evidence for a great effect on accent placement and the distribution of consonant clusters in a word which was brought about by changes in the morphological structure of a word.

Consider the characteristics of two stages of the development of Persian, those of Old and Modern Persian. The data come from Oranskij (1960), Mackenzie (1967) and Pisowicz (1985).

Old Persian has the following characteristics: (i) the word never coincides with a stem, (ii) accent falls on the penultimate syllable if it is heavy, otherwise accent falls on the antepenultimate syllable and (iii) obstruent clusters appear only word-initially and medially, never word-finally. The examples in (16) illustrate word-initial clusters.

(16) *Old Persian*

xʃáθra-	‘kingdom’
spádi-pati-	‘general’

Characteristics of Modern Persian are as follows: (i) the word coincides with the stem (lexical morpheme), (ii) accent falls on the final syllable and (iii) obstruent clusters are permitted in word-final and medial positions. No obstruent cluster is permitted in word-initial position. In loan words with word-initial clusters, a vowel epenthesises, either preceding or breaking up the clusters, e.g. Russian /stakan/ ‘glass’ is pronounced as [estakan] in Persian. The examples in (17) illustrate word-final consonant clusters.

(17) *Modern Persian*

deráxt	'tree'
lakpóʃt	'turtle'
dózd	'thief'

The following correspondences illustrate the changes that occurred during the development of Persian:

(18)

Old Persian		*Modern Persian*	
spádi-pati-	>	sipahbád	'general'
brátar-	>	baradár	'brother'

The deletion of the final suffixes (e.g. case markers) is correlated with a change of word accent placement (from the antepenultimate to the word-final syllable in Modern Persian (see (18)) and with the distribution of consonant clusters (from word-initial to word-final position (compare the examples in (16) and (17)). Thus, the Persian data suggest that the change of morphological type (from inflectional to isolated) is intertwined with changes in the prosody and phonotactics of the language.

Another universal characteristic of language is that it changes. This justifies the formulation of a language as an open and not a closed system. Languages are constantly in a state of flux; they are never fixed in any detail, but in each of them there are necessarily points that are liable to change even within a single generation. This is an inevitable consequence of the very essence of language and of the way in which it is handed down from one generation to the next (Jespersen 1924:31). The openness of a system explains the coexistence of regularities and irregularities in a language. Consequently, any linguistic theory will encounter irregularities which remain unexplained. There are some linguistic forms discussed in this thesis which, for different reasons, are unexplained, e.g. either because they are marginal within a certain language system or because their origin is unknown.

One aspect of language change concerns the phonotactic structure of a language, the central topic of this thesis. A variety of factors can play a role in the change of phonotactic structures, e.g. language-internal factors (morphological, phonological, etc.) and language-external factors (language contact).

In this thesis, I will suggest that Georgian has undergone a change from the CVCVCV type to CCC, mainly due to language-internal factors. Thus, morphological and phonological factors are argued to have played an important part in the transformation of the phonotactic structures of Georgian.

1.4. Methodology

Recognising that language is an open system means recognising that linguistic structures undergo continuous change. For ease of exposition, I will use a metaphor: "If we sit at a bank and look closely at the water, we predominantly see flow, i.e.,

process. If we step back we may be able to observe both flow and structure. If we fly in an airplane above the river we only observe structure." (Israel 1979:117). Traditionally in linguistics one chooses the airplane perspective of concentrating on structure. I take the second option and argue in this thesis that linguistic structures should be analysed by taking into consideration not only 'synchronic' patterns of a language, but 'diachronic'[10] ones too, as linguistic structures are the result of the process.

The types of evidence considered in the analysis of Georgian consonant sequences to be developed here, the Gradual Consonant Analysis (see Chapter 6), are as follows: (i) paradigmatic and syntagmatic, (ii) historical, (iii) phonetic and (iv) comparative. These types of evidence establish the generalisations of Georgian phonotactics and are subsequently incorporated in the analysis.

The analysis does not assume any underlying or a priori representations. The principles, which are assumed to be universal, are functional phonological principles united under an umbrella principle the *Balancing Principle.* From the conception of a language as an open system, it is predicted that structures undergo a change from 'simple' to 'complex' and vice versa. The reason of the change can both be language-internal and external.

The analysis starts with a careful investigation of immediately observable 'problematic' structures. Different types of evidence, e.g. syntagmatic, paradigmatic, phonetic, historical, comparative, etc., are taken into account to analyse the constituency and nature of such structures. At the final stage of the analysis, we arrive at abstract representations of the structures, theoretical constructs. Direct and clear links between the empirical structures and the theoretical constructs explain why the structures appear to be problematic. The analysis illustrates that *problem* in fact is a relative notion (see Chapter 6).

1.5. Proposals

The central claim of the thesis is that language is an open system. This bears directly on the methodology used in the analysis of Georgian consonant clusters, the Gradual Consonant Analysis. Unlike previous approaches to Georgian consonant sequences, e.g. Nepveu (1994) and Bush (1997), I argue that historical and comparative evidence are important for a proper analysis of the sequences. Other claims which will be made with respect to phonotactics in general, and specifically with respect to Georgian consonant clusters, are as follows:

[10] The opposition synchronic vs. diachronic is based on a dualistic standpoint and was introduced by Saussure (1967). I think no sharp distinction can be maintained between the two, as language continuously undergoes change.

(19) a) Phonotactics is dynamic.
b) Phonotactics is sensitive to morphology.
c) The stem is a phonological domain.
d) Consonantal complexity in Modern Georgian is due to morphological complexity, as well as the result of processes of vowel deletion and of complex segment formation.

1.6. Outline

In Chapter 2, I address the two main questions that arise with respect to the study of the phonotactics of a language: (i) where do phonotactic constraints hold and (ii) what are the principles or constraints governing co-occurrence restrictions? Thus, the domain of co-occurrence restrictions and the principles which govern them are discussed, in that order. Based on various phonological criteria, e.g. phonological processes, phonotactic restrictions, accent assignment, etc., I argue that besides the word, the stem (not the syllable, as suggested in previous studies) should be recognised as a phonological domain. Since the importance of the morphological constituency of the word is recognised, morphological typology is considered next. The following sections deal with the morphological classification of languages, which serves as a background for the discussion of phonological principles: the Obligatory Contour Principle (OCP), the Sonority Sequencing Principle (SSP) and the Principle of Resolvability (PR). The traditional formulation of the OCP is assumed and substantiated by additional language data. I propose a re-examination of the SSP by defining this principle at the stem/word domain, instead of the syllable domain, as was claimed in previous studies. In addition, unification of the SSP with a closely related constraint, the Syllable Contact Law (SCL), is suggested. In order to substantiate this proposal, the typological data provided by Zubkova (1990) are discussed. The generalisations illustrate that both the SSP and the SCL function within a stem/word domain and reflect the rising and falling sonority pattern of a word. Furthermore, the PR is re-examined and extended to two-member consonant clusters. The reconsideration of the principle leads to a hypothesis about the secondary nature of consonant clusters, which says that the patterns of the clusters are derived from the patterns of the stem. The hypothesis is tested in Chapter 3. Finally, it is suggested that the OCP, the SSP, the SCL and the PR are different instantiations of the Balancing Principle (BP). All the principles imply the universal principles of ease of articulation and perception. They are defined on a single domain, the stem or the word, depending on the language type. The principal message of this chapter is that the word has hierarchical structure, which includes the stem as a phonological domain occupying a place between the word domain and the segment. The structure has the shape of a pyramid. The pyramid representation of the word suggests that word phonotactics should be viewed simultaneously in three dimensions: meaning, form and structure.

Chapter 3 presents the main observations about the word-level phonotactics of Georgian. Georgian, as a Grammatical language (i.e. a language with features of inflectional and agglutinative morphology), demonstrates the discrepancy between lexical and grammatical morphemes in terms of its use of the phonological inventory. This observation presents an important clue for the study of the phonotactics of Georgian consonants. Several generalisations concerning the minimal word and the status of some consonant sequences are proposed. The disyllabicity of the Georgian minimal word is substantiated by phonological (accent assignment), phonetic (monosyllabic lengthening) and morphological (alienable/inalienable constructions) evidence. In this way, it is established that the minimal word in Georgian is of the $C_1V_1C_2V_2$ type. Characteristics of consonants both in isolation and in sequence are examined. I argue that Georgian sonorants are phonetically syllabic in consonant sequences. The claim is based on phonological, distributional, historical and comparative evidence. This claim especially concerns the most sonorant consonant, /r/. Convincing evidence of several kinds (phonological processes, distributional facts, historical considerations, etc.) suggests that harmonic clusters and the combinations of C + /v/ can be analysed as complex segments. The investigation of the word-initial consonant sequences shows that the complexity of consonant clusterings is illusory in Georgian. The most frequently realised clusters are combinations of obstruent + sonorant, harmonic clusters and consonants that share the laryngeal specification and are deccessive (i.e. [front] followed by [back] place of articulation). Similar consonant co-occurrence restrictions are found in two other contexts: across a vowel within a stem and in adjacency. The plausibility of the hypothesis proposed in Chapter 1 is thus confirmed by the Georgian data.

Chapter 4 presents a perceptual experiment substantiating the phonological status of Georgian harmonic clusters as complex segments. In Chapter 5, reduplication data from Georgian are presented and discussed to uncover the unmarked prosodic and segmental structure of the language. The reduplicative patterns support the claims made in Chapter 3: (i) The minimal word in Georgian is disyllabic; (ii) Harmonic clusters can be analysed as complex segments; (iii) C + /v/ combinations can be analysed as complex segments; (iv) /s/ + obstruent can be analysed as a complex segment; (v) Obstruent + sonorant is the most unmarked consonant cluster; and (vi) Georgian maximally allows two-member consonant clusters stem-initially. Consequently the data support the claim that consonant sequences are, in general, derived.

Chapter 6 examines consonantal verbal stems. Comparative and historical evidence suggests that Modern Georgian consonantal stems containing up to four consonants are derived from stems that have an epenthetic vowel between the consonants. The Gradual Consonant Analysis (GCA) is introduced in this chapter. The application of the GCA is demonstrated by an analysis of Georgian verbal and nominal forms containing long consonant sequences.

Chapter 7 presents some recent analyses of Georgian consonant clusters: the Headless Syllable Analysis (Nepveu 1994), the Syllabified Consonant Analysis (Bush 1997) and an analysis within the framework of Government Phonology (Toft 1999). Comments are made concerning specific aspects of a particular analysis. At

the end of the chapter a comparison is made between previous analyses and the GCA, and general conclusions are drawn.

In Chapter 8, a summary of the main findings and conclusions are provided.

2 Phonotactics: Domains and Principles

2.0. Introduction

The study of consonant clustering, the main theme of this thesis, is part of a broader field called phonotactics. I believe the specific issue of consonant clustering cannot be accounted for adequately without addressing general problems of phonotactics. Thus, the main goal of this chapter is to consider theoretical and methodological problems encountered in the analysis of the phonotactic structure of a language. In doing so, I introduce the reader to the concepts and tools used in the subsequent analysis. I will return to Georgian by presenting descriptive data (see Chapter 3) and an analysis, the Gradual Consonant Analysis (see Chapter 6).

Phonotactics is the study of the set of permissible arrangements or sequences of sounds in a given language. For instance, it is a well-acknowledged fact that not any combination of sounds can appear in a language, e.g. clusters such as *bk, *gs, *p'g, are ill-formed in Georgian, as in many other languages. Co-occurrence restrictions are language-specific, e.g. a word beginning with the consonant cluster *zn* violates the phonotactics of English, but not of Georgian. On the other hand, as cross-linguistic studies of phonotactic patterns have shown, languages do have certain phenomena in common. This observation made possible the formulation of a number of phonotactic universals attested across languages of the world (Trnka 1936, Trubetzkoy 1939, Saporta 1963 and Greenberg 1978, among others). I consider the characteristics of both universal and language-particular constraints later in this chapter.

Two questions that arise with respect to the phonotactic constraints of a given language are *where?* and *how?*; i.e. a researcher interested in phonotactics addresses the following two problems:

(1) a) Where do phonotactic constraints hold?
b) What are the principles or constraints governing co-occurrence restrictions?

Thus, both the domain of co-occurrence restrictions and the principles which govern them are important. The following discussion addresses these questions, in the order in (1).

The main goal of this chapter is to propose a phonological hierarchy which includes the stem as a phonological domain occupying a position between the word and the segment. Thus three main units, the segment, the stem and the word, are argued to be necessary to account for the phonotactic structure of a word. With

respect to the issue of constraints I propose that various phonological principles, the Sonority Sequencing Principle, the Obligatory Contour Principle and the Syllable Contact Law, are different instantiations of the Balancing Principle, whose domain is the stem or the word, depending on the language type.

2.1. Domains

2.1.0. Introduction

Traditionally, evidence from several areas has been used to define the phonological domain.

(2) a) Phonological rules
b) Phonotactic generalisations
c) Minimality constraints
d) Prosody (accent or tone assignment)

Proper domains have been established by taking into account different aspects of the phonological organisation. In this thesis I concentrate primarily on the phonotactic aspect. This does not mean, though, that the domains I argue for are inadequate for stating phonological rules or prosodic patterns. They are considered too, but less extensively.

One of the central preoccupations of phonological theory relevant to phonotactic analysis, in particular to the analysis of consonant clusters, is the determination of a proper domain, or in representational theories, a level of representation.

It should be emphasised from the outset that in this thesis the study of phonotactics, and specifically of consonant sequences, is restricted to the word domain. This is due to the fact that the word, being a relatively independent unit of language, is the most uncontroversial domain for stating phonotactic generalisations (Booij 1999, Ewen & van der Hulst 2001). Derived words and phrasal structures do not fall within the main scope of this thesis. However, throughout the thesis, derivation (whether by suffixation or prefixation) is argued to be one of the sources of consonant clustering. The word domain in Georgian includes the stem and an optional suffix, e.g. the nominative case marker /-i/ in nominal forms. Since the domain of the study has been defined, further clarification of the terminology is of relevance.

The definition of *word* is far from trivial. The word is the focus of the multidimensional interaction of syntax, morphology, phonology and semantics, and since the realisation of this interaction varies from language to language, a uniform and coherent definition of the word is almost impossible to formulate. In addition, the definition of a word ultimately mirrors the researcher's methodological-theoretical bias, which is based on a certain understanding of the organisation and functioning of the language faculty.

2.1.1. The word

One way to try to define *word-hood* is in terms of one of the linguistic modules, e.g. phonology, syntax or morphology. This is the strategy followed in most contemporary linguistic theories, where the term word is specified for certain sub-domains. For instance, *phonological word*, *prosodic word*, *morphological word*, etc. are different entities for phonological processes and accent assignment (see Booij 1983).

Stemming from the view that a linguistic sign consists of meaning and form as indivisible components (see Chapter 1), the term *word* in this thesis has a uniform definition. The word is a minimal independent unit, well-formed both in terms of phonology/prosody and in terms of morphology. Since this thesis does not deal with derived words, the term *word* used throughout is close to the notion of *phonological word*.

Narrowing down the study of phonotactics to word domain begs certain questions bearing on the issue of sub-constituency at word level.

(3) a) Is the word domain a necessary and/or sufficient unit for accounting for word-level phonotactics?
b) Do we require recourse to additional representational levels below word level? And if so, which ones are important?

As I proceed to answer these questions, I will attempt to show that establishing the nature of sub-constituency is an important precursor to the proper description, explanation and analysis of word-level phonotactics.

Before addressing the issue of sub-constituents of the word level, I consider some empirical observations illustrating the word as a domain of phonotactic generalisations.[1]

An often-cited example is that of Italian. Italian words are subject to a prosodic minimality constraint (Thornton 1996), in that they must be minimally disyllabic. Italian words usually end in a vowel which functions as a morphological ending, e.g. the lexical morpheme *pizz-* never occurs independently, but is always followed by the ending /-a/ or /-e/, forming e.g. *pizza* ‘pizza’. The same can be said about Georgian: words always end in a vowel, e.g. the lexical morphemes *kac-* and *c’ign-* never occur independently, but the nominative suffix /-i/ is added to form the well-formed words *kaci* ‘man’ and *c’igni* ‘book’.

[1] There are numerous studies arguing that the word is a domain for other reasons as well, such as: (i) Phonological rules; e.g. Hungarian vowel harmony applies only when its trigger and target belong to the same prosodic word (Booij 1984, Nespor & Vogel 1986); (ii) Minimality constraints; for instance, in Australian languages words can typically be no smaller than two syllables (Dixon 1980). In other languages the restrictions governing minimal size involve not the syllable, but the mora. For example, in the Cushitic language Iraqw the smallest prosodic word is bimoraic, i.e. prosodic words of the form V: and VC are allowed because these sequences contain two moras, whereas those consisting of a single short vowel, e.g. CV, are not (Mous 1993:26, Hall 1999:8); (iii) Studies reporting the word as the domain of stress or tone assignment are too numerous to list here (but see Dixon 1977, Nespor & Vogel 1986, Raffelsiefen 1999, Russell 1999).

Consider also well-known distributional generalisations referring to word-initial and word-final positions: e.g. /h/ in Georgian can occur only in word-initial position, never in word-final position; /ŋ/ in English occurs in word-final position, not in word-initial position. In Bare, a language belonging to the Northern Arawak languages of the Brazilian Upper Rio Negro, consonant aspiration occurs only in word-initial position, e.g. /hmi/ 'hammock', /phani/ 'house' and /hwiyuɻi/ 'prey'. However, vowel nasalisation occurs only in word-final position, e.g. /kuhũ/ 'he', /titĩ/ 'a kind of monkey' (Aikhenvald 1996).

Some phonotactic generalisations concerning co-occurrence restrictions, e.g. the Syllable Contact Law (SCL) and the Obligatory Contour Principle (OCP), are argued to be operative within the word domain as well (see Booij 1999).

There are interesting cases of phonotactic generalisations in languages where a lexical morpheme formally coincides with a word. In such languages it is difficult to distinguish whether the generalisations are characteristic of a word domain or of a stem (lexical morpheme) domain. Consider for instance, phonotactic restrictions on monomorphemic words of English. The generalisations are taken from Davis (1985).[2]

(4) Words of the type $C_iC_jVC_j$ are not attested. Thus, forms like **flil* are not possible.

Similarly:

(5) $C_iC_jVC_jC_i$ forms, e.g. forms like **flilf*, are not attested.

While the phonotactic patterns of words in Italian and Georgian mentioned above are exclusively characteristic of the word domain, the patterns presented in (4) and (5), due to the formal coincidence of a word with a lexical morpheme (i.e. stem) and a syllable, could be viewed as characteristics of either a word, a stem or a syllable. It should be mentioned from the outset that the generalisations in (4) and (5), due to the formal match of domains, cannot be used as convincing evidence to argue for the independent status of units below word level.

Having established the word as one of the domains of phonotactic generalisations, the problem now is to investigate whether there are smaller units/constituents involved in phonotactic restrictions. Consider two generalisations where phonotactic restrictions seem to refer to different domains: (i) the word and the syllable and (ii) the word and lexical morpheme (stem).[3] The following two generalisations capture the distribution of the high vowels /i/ and /u/ in Georgian.

[2] Davis (1985) discusses these generalisations to demonstrate that there appear to be systematic phonotactic constraints between onset and coda. Moreover, these empirical data are used to argue for a 'flat' structure of the syllable.

[3] In this thesis, the term *stem* is used interchangeably with the term *lexical morpheme*.

(6) The vowel /u/ is commonly found in the final position of monosyllabic words, e.g. /bu/ ‘owl’, /ru/ ‘brook’, /χ’ru/ ‘deaf’; however, its occurrence in the final position of polysyllabic words is very restricted (an exception to the latter is /juju/ ‘breast’ (reduplicated form)).

(7) There is no stem (lexical morpheme) ending in the vowel /i/,[4] however, most words in Georgian end in the vowel /i/, e.g. /xeli/ ‘hand’, /c’erili/ ‘letter’ and /p’uri/ ‘bread’.

In line with (7) we find a generalisation about the distribution of consonant clusters in Georgian.

(8) Consonant clusters can appear in both stem-initial and stem-final positions, e.g. /cxvir-i/ ‘nose’, /marcχ’v-i/ ‘strawberry’. Consonant clusters occur only in word-initial position, never in word-final position, e.g. /brjola/ ‘fight’, /brč’χ’ali/ ‘claw’.

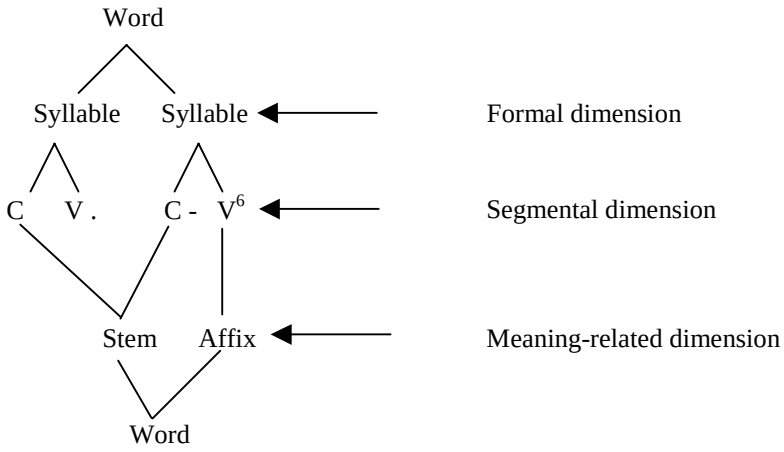
The generalisations in (6), (7) and (8) illustrate that more levels, in addition to the word level, are necessary for phonotactic generalisations. In the case of (6), the distribution of the vowel /u/ is sensitive to the number of syllables in a word. In (7) and (8), the generalisations are formulated separately for the stem domain and the word domain. The generalisation given in (6) is different from those in (7) and (8). Both the restrictions in (6) are characteristic exclusively of the word domain, either monosyllabic or polysyllabic. ‘Syllable’ in these generalisations is used as a short-hand. However, the restrictions in (7) and (8) are stated separately for the word and stem domains. I will return to the discussion of the difference between these two types of generalisations in following sub-sections (see 2.1.2 and 2.1.3).

Since the stem and the syllable are fundamentally different units, the incorporation of both in an analysis of word-level phonotactics presupposes viewing a word simultaneously from different angles, or different dimensions. To illustrate this, the representation of an arbitrary CVCV word structure in terms of formal and meaning-related (i.e. morphological) components is given in (9) (the two components are represented on separate planes).[5]

[4] The two loan words /čai/ ‘tea’ and /t’ramvai/ ‘tram’ are exceptions to this generalisation. A stem in general can end in any other vowel: /a e u o/, e.g. /da/ ‘sister’, /xe/ ‘tree’, /k’u/ ‘turtle’, /xbo/ ‘calf’, etc.

[5] The names of dimensions at this stage are not very accurate and are given mostly to indicate that there are three different aspects involved in word-level phonotactics. A revised version is offered in (96) in this chapter.

(9)



For ease of exposition I have chosen a type of CVCV word which is characteristic of languages like Italian, Georgian, Sanskrit, Old Persian and Russian, i.e. one that contains an inflectional suffix, in this case the word-final -V in the CVC-V word. In such languages the suffix usually denotes a nominative case, gender, determiner or class marker. These languages, already having a lexical and a grammatical morpheme in the minimal word-form, illustrate the asymmetry between lexical and grammatical morphemes and their respective phonotactic patterns. In addition, as a sub-constituent of a word, the stem is characterised by its potential autonomy from the word. This does not mean, however that languages which are characterised by the formal coincidence of a word with a lexical morpheme, e.g. Dutch, English, Modern Persian, do not make reference to the lexical vs. grammatical asymmetry, and the same holds for languages that have the following type of structure: obligatory prefix + stem, e.g. some of the Bantu languages. Regardless of the formal coincidence of a stem and a word in languages such as English or Dutch, the stem (lexical morpheme) functions as an independent domain from a word (see arguments discussed by Booij 2000). The same applies to some of the Bantu languages having the word structure prefix + stem (see Downing 1999).

Thus far I have illustrated that there are constituents smaller than the word, which can serve as domains for phonotactic generalisations. In the remainder of this chapter I will investigate the plausibility of studying word phonotactics from the perspectives of form (see section 2.1.2) and meaning (see 2.1.3). Finally, I suggest a unified approach which incorporates the insights of both.

It is plausible that lower units should meet two main requirements in order to have the status of a phonotactic (i.e. phonological) domain.

[6] The dot refers to a syllable boundary and the dash to a morphological boundary.

(10) a) The unit should be a part of the formal side of a word. Thus, it should be a sub-constituent.[7]
b) The unit should function independently from a word, e.g. it should have distinct phonotactic patterns.[8]

Both the syllable and the stem meet the requirements formulated in (10a): both are sub-constituents of the word. What remains to be demonstrated is whether these units are independent from the word domain, and in which respects. In the following sub-sections, I examine the syllable and the stem with respect to the latter criterion.[9]

2.1.2. The syllable

2.1.2.0. Introduction

It is remarkable that both proponents and opponents of the syllable refer to the syllable with a certain amount of pessimism.

"The numerous attempts to give a syllable a rigorous phonetic and/or phonological definition (Stetson 1951, Hjelmslev 1939, Haugen 1956, Hála 1961, Rosetti 1962, O'Connor & Trim 1953 and Pulgram 1970) have not been entirely successful in accounting for the wide range of data associated with the syllable" (Bell & Hooper 1978:4).

An opponent of the syllable notes that "the term can be formally taught as a means of labelling some aspect of phonological reality, but it is by no means always obvious exactly what that reality is. In alphabetic traditions, calculating the number of syllables in a word typically involves identifying the number of peaks of perceptual prominence, each typically associated with a vowel" (Harris 1994:45).

The definition of the syllable is usually based on the inherent sonority of phonological strings (see Spencer 1996[10] and Blevins 1995, among others): "… the syllable is the phonological unit which organises segmental melodies in terms of sonority; syllabic segments are equivalent to sonority peaks within these organisational units" (Blevins 1995:207). The fact that no current phonetic study can provide

---

[7] Note that unlike in Prosodic Phonology (Nespor & Vogel 1986) lower units are not required to be exhaustively part of upper units (i.e. words in this case).

[8] Note that phonotactic patterns together with phonological rules, and that rules of accent assignment are necessary criteria for defining the domain.

[9] I do not discuss the constituent *foot* in this chapter or in the rest of the thesis. The importance of this unit was originally proposed in order to account for accent placement in a language (see among others Liberman & Prince 1977, Halle & Vergnaud 1978). Nespor & Vogel (1986) advocate the *foot* as a domain of phonotactic generalisations, but I believe that in all cases, e.g. aspiration in English and phonotactic restrictions in Ž|'hõasi, a Namibian language, the generalisations can be stated in terms of a stressed syllable or a stem. Selkirk claims that the foot does not exist as a constituent in phonology at all since "most alleged foot sensitive rules can be recast as rules sensitive to the stressed-stressless distinction" Selkirk (1984b:31). Besides, I am not aware of any word-level phonotactic generalisation referring exclusively to the foot domain. Taken all these into consideration, I do not go into details of argumentation here, but simply ignore the foot as a possible domain of word-level phonotactics.

[10] As Spencer (1996) notes, "syllable structure can be determined just from the segmental composition of a word" (Spencer 1996:96).

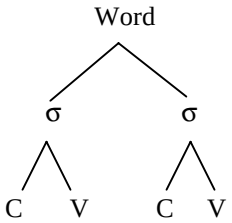
a uniform determination of a sonority scale leads to the conclusion that the phonetically defined phonological syllable has little phonetic support.[11]

Due to the impressionistic definition of the syllable, many researchers have ignored the syllable in their phonological descriptions of a language (Chomsky & Halle 1968, Hyman 1985, Kaye, Lowenstamm & Vergnaud 1985, Harris 1994).

Anderson (1969), Fudge (1969), Vennemann (1972) and Hooper (1972) are among those who recognise the importance of the syllable in phonology. In current phonological theories much work has been devoted to the problem of syllable-internal constituency (Pike & Pike 1947, Kuryłowicz 1948, Fudge 1969, Kahn 1976, Selkirk 1982, Clements & Keyser 1983, Hyman 1985, McCarthy & Prince 1986 and Hayes 1989). The following discussion will show that the importance of syllable as a domain is not recognised in this thesis. For this reason, the issue of syllable-internal constituency falls outside the scope of the thesis.

Since the syllable constitutes a formal aspect of a word, for ease of exposition I consider a part of the representation (9), repeated in (11).

(11)



As seen from the schema in (11), there are two potential threats to the autonomous status of the syllable.

(12) a) The word CVCV is subdivided into two syllables. The word is exhaustively syllabified. Exhaustiveness, as argued in 2.1.2.1 and 2.1.2.2, is the main reason why the syllable cannot function independently from a word.

b) Another reason that inhibits the independence of a syllable as a domain is related to its dependence on inherent sonority.[12] It is the artefact of rising and falling sonority patterns attested in speech production. In terms of Haugen's definition, the syllable is "… the smallest unit of recurrent phonemic sequences" (Haugen 1956:216).

[11] For the difference between the phonetic syllable and phonological syllable see Pike (1947), Fischer-Jørgensen (1952) and Hooper (1972).

[12] The connection between the syllable and sonority has led many researchers to correlate the syllable with the Sonority Sequencing Principle (hereafter, the SSP). Later, in section 3 of this chapter, I demonstrate that the SSP can be also defined on the stem, or word domain. Both are also the domains of other principles of phonotactic organisation, e.g. the Obligatory Contour Principle.

Both observations are interrelated and lead to the assumption that what is "characteristic" to a syllable is actually derivable either from the word or from the internal constituency of a segment.

Leaving potential threats aside for the time being, let us consider several arguments which have been proposed to justify the introduction of the syllable into current phonological theories. It has been argued in many instances that there are several types of phonological generalisations that necessarily require direct reference to the syllable. Some of these are as follows (see Hooper 1972, Kenstowicz 1994, Blevins 1995).

(13) a) Syllabification
b) Phonotactics
c) Word-edge phenomena
d) Phonological processes
e) Accent assignment

I proceed with the examination of the evidence for each of these in the subsequent sections.

2.1.2.1. Syllabification

One way of establishing the independence of a unit involves identifying its boundaries in an unambiguous fashion. Hence, extensive work has been devoted to the issue of syllabification and has been used as evidence for the syllable in phonological theory.

I proceed with arguments concerning the validity of the unit *syllable* in phonology, and especially for syllabification and phonotactics, as suggested by Kohler (1966):

"If the syllable has any real status in phonology, its boundaries must be discernible. A grammatical formative in any language can be rewritten as a sequence of elements like

$$GF \rightarrow C_1VCVC\ldots CVC_2$$

Where C_1 stands for any non-arbitrary pre-vocalic consonant (cluster), C_2 for any non-arbitrary post-vocalic consonant (cluster) and C for any consonant (cluster) between two vowels. Any part right of the arrow may be zero.

C can now be rewritten as

$$C \rightarrow C_1, C_2, C_3, C_2\,C_1$$

$(C \rightarrow C_1\,C_2$ is impossible by definition)

Argument 1. If $C_1 \neq C_2$ and if $C \rightarrow C_1, C_2, C_1\,C_2$, a non-arbitrary syllable division is possible but is implicit in the definitions for C_1 and C_2" (Kohler 1966:207).

Consider a case of syllable division of a CVCCV string in Spanish and English, as discussed in Hooper (1972). When two or more non-syllabic segments co-occur, there are several possibilities for syllable division, depending on the feature composition of these segments. First, if there are two non-sonorant segments, the syllable boundary will be inserted between them, e.g. Sp. *ap.to* 'apt', Eng.

ab.solute. Second, if there is one non-sonorant (not followed by a nasal), the syllable boundary is inserted before it, no matter what the other segments are; Sp. *ar.te* 'art', *pa.dre* 'father', *con.tra* 'against', *a.byerto* 'open', Eng. *con.tent*. Thus, obstruents followed by glides or liquids act as syllable onsets.

However, there are several exceptions to this rule. In most languages that allow clusters of obstruents and liquids, coronal obstruents do not precede /l/, i.e. /tl/ and /dl/ are not allowed initially. When these segments co-occur within a word, it is common for the syllable boundary to separate them, e.g. Eng. *At.lan.tic* vs. *a.tro.cious* (Hooper 1972). The difference in syllabification of /tl/ vs. /tr/ quite clearly follows from the fact that English allows word-initial /tr/ clusters, while /tl/ is not a possible cluster in English.

Consider similar examples from Spanish and English, as discussed in Hooper (1972). The place of the syllable boundary in relation to /s/ varies from language to language and thus cannot be generally predicted. For instance, the previous rule will place the syllable boundary between the /s/ and a following obstruent if the cluster is flanked by vowels: /Vs.tV/. So the correct results are generated for Spanish *es.ter*, *es.pecial*; but not for English, where the syllable division appears before the /s/: *a.sparagus*. Once again the restrictions on syllable-initial clusters seem to be the same as those on word-initial clusters. Spanish does not allow clusters of /s/ + obstruent initially, but English does.

These patterns of syllabification show that the syllabification of underived words cannot be determined straightforwardly unless word initial/final patterns are taken into account.

Argument 2. "If $C_1 = C_2$ there are cases where a syllable division is not determinable, and the syllable is therefore an impossible concept. The same applies to C $\rightarrow C_3$, i.e. a consonant (cluster) that can only occur intervocalically and, therefore, not be uniquely determined as pre- or post-vocalic" (Kohler 1966:207). Consider the English word *master*. According to the principle of Maximal Open Syllabicity (Pulgram 1970), it is syllabified as *ma.ster*. Notice, however, that the syllabification creates a sequence which violates a sequential constraint in English by which the lax vowels /ɪ ɛ ʊ ɔ æ/ are disallowed in word-final position. Even more indeterminate is the syllabification of the Spanish word *transkriber*. The syllabification *tran.skriber* creates the initial *skr* cluster, which is disallowed word-initially in Spanish, while the syllabification *trans.kriber* creates a final *ns* cluster which is unacceptable word-finally in Spanish.

There are also cases when the indeterminacy of a syllable boundary is caused by two possible syllabifications in a language. Consider a case in Polish: "the favoured syllabification of *dobry* 'good' and *pat[sʲ]eć* 'look' are … *do.bry* and *pa.t[sʲ]eć*, though *dob.ry* and *pat.[sʲ]eć* are possible as well" (Kenstowicz 1994:262).

Argument 3. "If in any language the division into syllables clashes with morphemic divisions the concept of a syllable is harmful" (Kohler 1966:207).

Consider a case discussed in Ewen & van der Hulst (2001). In some cases, morpho-syntactic boundaries may prevent the application of the 'normal' phonological syllabification rules of a language. In an intervocalic biconsonantal cluster in

English, the syllable boundary falls before the first consonant. The same pattern occurs in Dutch, as shown in (14).

(14)	sterker	STERK+ER	'stronger'	$[stɛr]_σ$ $[kər]_σ$
	avontuur	AVONTUUR	'adventure'	$[a]_σ$ $[vɔn]_σ$ $[tyːr]_σ$

However, if the morphological boundary is strong enough, syllabification does not operate in the expected way, as shown in (15).

(15)	avonduur	AVOND#UUR	'evening (hour)'	$[a]_σ$ $[vɔnt]_σ$ $[yːr]_σ$

Another case when syllabification is determined not by syllable-internal principles, but by the stress pattern is the following: in American English /sentər/ 'centre' becomes [senr;ˌ], while /séntɔ̀r/ 'centaur' remains [séntɔ̀r]. "A possible interpretation of this is that the difference in the stress pattern entails a difference in the syllabification of these words" (Vennemann 1972:17).

Lack of evidence for the autonomy of a syllable as a unit in syllabification has led many to say that it is unnecessary concept and should only be used as a quick reference in cases of the non-arbitrary syllable division (to state the Syllable Contact Law) (Kohler 1966). In this chapter, in section 2.2.3, I will demonstrate that the Syllable Contact Law and the Sonority Sequencing Principle are different instantiations of a single phenomenon and characteristics not of the syllable, but of the stem and the word.

In order to sustain the notion of the syllable as a level of representation, the following questions must be answered.

(16) a) At which level should syllabification be defined?
b) Where do syllables come from?
c) Are they present in the lexicon or are they generated in the course of the phonological derivation?

The final question is the most crucial for proposals made in this thesis. Blevins (1995) claims that "three observations suggest that in the general case, syllable structure is not present in underlying representations: (i) minimal pairs distinguished by syllabification alone are rare, and are nonexistent in many languages; (ii) segments in many languages exhibit syllabicity alternations which can be viewed as the simple result of derived syllabification; (iii) individual morphemes often fail to conform to the possible syllable types of a given language, making lexical syllabification infelicitous" (Blevins 1995:221). The claim is in line with the assumption that syllable structure is largely absent from lexical representations and is either constructed or mapped by rule (e.g. Vennemann 1972, Kahn 1976, Levin 1985 and Itô 1986) or supplied by a phonological generator (e.g. Prince & Smolensky 1993). For a different view, see Harris & Gussmann (1998).

The absence of the syllable from the lexicon is supported by numerous studies on language production (e.g. Levelt 1992, Meijer 1994, 1996 and Roelofs 1996). Schiller (1997) studied the syllable's role as a processing unit in Dutch speech production. Masked syllable priming experiments and syllable reversal experiments substantiated previous claims that syllable units are not represented in the lexicon but emerge during prosodification (the same results were obtained by Baumann 1995). The conceptual background to these experiments is the two-stage model of lexical access (Kempen & Huijbers 1983, Levelt & Maasen 1981, Levelt 1992), which assumes that the phonological encoder generates a sound form for each lemma and for the utterance as a whole (Levelt 1992). Two kinds of phonological information are distinguished: segmental and metrical. Segmental and metrical information are supposed to be retrieved separately (Meijer 1994, 1996 and Roelofs & Meyer in press). "Spelled-out segments are combined with these metrical frames in a process called segment-to-frame association. The model assumes that segments are not marked for syllable position but just for their serial position within a morpheme (Roelofs 1996)" Schiller (1997:2). The output of phonological encoding is a phonological word, which is fully specified with respect to its metrical and syllabic structure.[13]

I accept the insights of these studies and in the remainder of this thesis I argue for the irrelevance of the syllable, and consequently, the syllable domain in the lexicon.

In addition, researchers who wish to maintain the notion 'syllable' need to distinguish between two syllable types (related to the general conception of the two-stage approach, or, if one wishes, to a two-representation view: underlying vs. surface). For instance, Fudge (1969) defined two concepts of the syllable, one at an abstract level, which is not defined in phonetic terms, and one at a concrete level, which is defined in phonetic terms. The same insight is found in Brown (1969): "the syllable we are concerned with here is not, then, an underlying distributional unit, a phonological syllable, but the distributional unit of the phonetic realization" (Brown 1969:6). In word-production studies (e.g. in the two-stage model of lexical access) two types of syllables are also distinguished: phonetic and phonological. The phonetic syllables are conceived of as articulatory motor units, i.e. execution units, while phonological syllables are regarded as planning units.[14] Schiller translates the difference into a dichotomy: lexeme syllable vs. speech syllable. In my approach, what is called in these cases phonological syllable or lexeme syllable is called a stem (see section 2.3).

[13] Besides production experiments, perception experiments on syllable monitoring also do not substantiate the claim that the syllable is the basic perceptual unit (Norris & Cutler 1988). The results of gating experiments could also be taken as evidence to strengthen this position (see Sawusch & Jusczyk 1981).

[14] Compare the discussion on whether the syllable may be a unit of performance rather than of competence (Fromkin 1968).

2.1.2.2. Phonotactics

Convincing evidence illustrating the independence of the syllabic domain would be the existence of languages with restrictions on the syllable that do not hold for the word. Since I do not know of any study illustrating the existence of such a correlation, consider the reverse implication advocated by Booij (1983): "… languages have special restrictions for the word-initial and word-final positions that are not valid for every syllable onset and coda, respectively". Some examples are the following:

(17)

Word-final restrictions		Source
Japanese	no word-final consonants	Vennemann (1978)
Huichol		Bell (1976)
Timicua		Granberry (1956)
Campa Arawak		Dirks (1953)
Sierra Nahuat	no word-final [m]	Key & Key (1953)
Arekuna Carib	no word-final [m], [n]	Edwards (1978)
Spanish	no word-final [ns]	Pulgram (1970)

Booij (1983) claims that in all these examples the restrictions apply only word-finally, but not syllable-finally.

Booij considers additional data from languages where restrictions hold word-initially, but not syllable-initially.

(18)

Word-initial restrictions		*Source*
Tamil	no word-initial lateral segments	Fowler (1954)
English	no word-initial [ž]	Hooper (1976:197)
Marihuana	no word-initial consonant clusters	Pike & Scott (1962)
Sierra Nahuat	no word-initial [h], [g]	Key & Key (1953)
Ongolo	no word-initial consonants (except certain prefixes)	Dixon (1970)
Dutch	no word-initial [pj], [kj]	Booij (1981)
Bamileke	no word-initial glottal stop	Hyman (1978)

I cannot in detail evaluate the importance of each of the generalisations formulated in (17) and (18) with respect to the syllable as an independent sub-constituent of a word, as argued in this thesis. I am arguing for the independent status of the syllable in a word, where the 'word' means a non-derived word. Some generalisations formulated in Booij (1983) are inconclusive for the present discussion, because some of the characteristic patterns of the syllable emerge in a process of a derivation. For example English does not have word-initial /ž/, but it occurs as an allophone of /z/ in derived environments, across word boundaries, e.g. *he sees you* [hiy.siy.žə] (Hyman 1978). The same applies to Dutch, with syllable-initial /pj/, /kj/ which occur only in derived environments. Hence, the Dutch and English cases do not contradict the claim that the syllable is not an independent sub-constituent of an underived word.

Some of the generalisations seem very marginal to certain languages and consequently cannot serve as strong arguments for positing a syllable as an independent domain of phonotactic generalisations. Consider, for example, the case of Arekuna Carib, which does not have word-final nasals /m/ and /n/. As shown in (17), the generalisation is formulated in phonetic terms. Indeed, as Edwards (1978) reports, the language has the alveolar nasal /n/ in word-final position, but this is realised as an allophone [ŋ], e.g. /aton/ [atɔŋ] 'cold'. This alternation [n]/[ŋ] in word-final position seems to accord well with the fact that only velars can occur word-finally: "word-finally only [k] and [ŋ] are possible" (Edwards 1978:226).[15] Note that both allophones of the nasal /n/ are attested in word-medial position: [ŋ] occurs when followed by a velar, elsewhere [n] occurs, e.g. [imaŋka] 'to fetch', [tano] 'big', [yenu] 'eye' and [ipanta] 'bench'. I do not see the necessity of formulating all these generalisations in terms of both the word and the syllable since this introduces a seemingly needless duplication in the grammar. The word, by distinguishing between word-initial, medial and final positions, will suffice. The syllable may yet be retained as a heuristic 'quick reference' tool for the linguist. Note additionally that syllable constituency cannot account or explain the alternation [n]/[ŋ] in medial position (both [i.maŋ.ka] and [i.pan.ta] are syllabified the same way: in both cases [n] as well as [ŋ] are in syllable-coda position), unless the segmental constituency of the string is taken into account; the alternation [n]/[ŋ] depends on the place of articulation of a following consonant, regardless of syllabification patterns.

Consider now the case of Sierra Nahuat, for which it has been claimed that it does not have word-final /m/ but does have syllable-final /m/. The evidence which would substantiate the generalisation would be word-medial clusters of /m/ + obstruent type, but as Key & Key (1953) report, while there are many word-medial sonorant + obstruent or obstruent + obstruent clusters in the language, there is only one type of a cluster beginning with the sonorant /m/, the /mp/ cluster. In the article, I found only one loan word which had the word-medial /mp/ cluster: *trumpet* (borrowed from Spanish). If the word-medial cluster /mp/ is truly marginal in the language, then the generalisation that the language has syllable-final /m/ lacks empirical grounding.

The generalisations I have discussed thus far merely refer to the occurrence of a certain phoneme or an allophone. Moreover phonotactics is the study of the co-occurrence restrictions in a language. Many co-occurrence restrictions hold between consonants across a vowel, which cannot be stated in terms of the syllable. Consider a hypothetical C_1VC_2V word, which has restrictions on the co-occurrence of C_1 and C_2. Traditionally, a C_1VC_2V word is divided into two syllables $C_1V.C_2V$, leaving C_1 and C_2 in different syllabic constituents. Unless the whole string C_1VC_2 or C_1VC_2V is recognised as a domain, these types of generalisations cannot be formulated. This leads us to the postulation of a stem domain. I return to this issue with additional arguments in section 2.2.3.

[15] The fact that velars are typical word-final consonants is substantiated in numerous studies (see section 2.4).

Blevins, while arguing for the syllable as a phonological domain, notes in a footnote, that: "surprisingly, there are few if any feature co-occurrence constraints which appear to take the syllable as their domain. Constraints in Sanskrit hold within the morphological stem; likewise, proposed constraints on the feature [+constricted glottis] in Mayan languages within the root; in Chinese languages, labial constraints have been claimed to hold within the syllable, however, this can also be viewed as a morpheme structure constraint as the syllable is equivalent to a morpheme. In sum, while syllable-internal features could provide potential evidence for the syllable as a phonological constituent, there are few convincing cases of such constraints" (Blevins 1995:235).

Studies on speech comprehension have substantiated the claim that phonotactic effects are derived from the lexicon (McClelland & Elmen 1986). More recent studies (Pitt & McQueen 1998, Vitevich & Luce 1998, 1999) suggest that phonotactic knowledge is stored (and used) at an earlier, pre-lexical stage of processing.

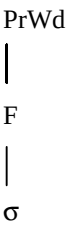
In section 2.3, additional evidence is provided for the lexical morpheme (stem) as the most optimal domain of phonotactic generalisation in numerous languages.

2.1.2.3. Word-edge phenomena

It is a well-known empirical fact that many languages allow segmental complexities at word edges. Proponents of the syllable, claiming that words are exhaustively parsed into well-formed syllables, find it problematic to explain and formalise such complex structures. For full-fledged arguments see section 2.3 and Chapter 6.

Many studies which attempt to analyse word-edge phenomena adopt insights from Metrical Phonology, introduced briefly below. The theory of Metrical Phonology proposes the Prosodic Hierarchy depicted in (19).

(19) *Prosodic Hierarchy* (partial)[16]



The Prosodic Hierarchy is constructed in accordance with various principles, one of which states:

[16] Since the discussion concerns the lower sub-constituents of a word I have given here only a sub-part of the original hierarchy.

(20) *The principle of exhaustive parsing*
A unit of a given level is exhaustively contained in the superordinate unit of which it is part (Nespor & Vogel 1986:7).

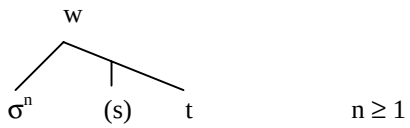
A restriction with the same effect is formulated under the name of "Strict Layer Hypothesis" (Selkirk 1984b), and later as the Prosodic Licensing Principle (Itô 1989).

(21) *Prosodic Licensing (PL)*
All phonological units must be prosodically licensed, that is, belong to higher prosodic structure.

The principle means that each unit should belong to a member of the next higher level in the prosodic hierarchy; segments must belong to syllables, syllables to metrical feet, and feet to prosodic words.

Having introduced the theoretical background, I turn to the empirical facts. Consider the case of Dutch. Dutch has a few words ending with three or four consonants, for example, *markt* 'market', *links* 'left' and *herfst* 'autumn'. Booij (1983) advocates that one of the dimensions of the syllable is that of the appendix,[17] and subsequently assumes the following canonical structure of the phonological word:

(22)



Note that positing an appendix associated directly to the word node violates the Prosodic Licensing Principle. Moreover, it is interesting that the appendix always consists of a coronal. This straightforwardly correlates with the fact that in Dutch consonantal suffixes always consist of one or more coronal obstruents, i.e. /t/, /s/ or /st/. In addition, there is the asymmetry between the left word edge and the right word edge. Consonant complexity is characteristic for the right edge, not for the left. This correlates with the fact that Dutch has no prefixes that consist of consonants only. This is not the case with Dutch suffixes. The phonotactic constraints on prefix/suffix seem to be related to word-edge constraints in Dutch. Similar correlations are discussed in section 2.3, in Chapter 3 and in Chapter 6. Thus, the consonantal complexity in Dutch is related to the morphological complexity of a word. It has to be admitted, though, that diachronically complex words are sometimes difficult to decompose for various reasons. Sometimes a grammatical affix is no longer productive and lexicalises together with the lexical morpheme. Another possible scenario is that a language borrows a morphologically complex word and treats it as a simplex one.

[17] Other languages for which appendices can be assumed are English, German Malayalam, Swedish, Berber and Zoque (Booij 1983). It would be interesting to study whether the presence of appendices in these languages is related to the morphological structure of a word.

In any case, the phonological complexity assumed can be explained, and is directly related to morphological composition, not to the syllable.

2.1.2.4. Phonological processes

Proponents of the syllable often argue for its usefulness in stating phonological rules. It is worth spelling out some of the arguments in full. Consider a process discussed in Vennemann (1972): "Let me first show you a subtle process which is in perfect harmony with the *Law of initials.*[18] Gothic has a well-known rule named Sievers' Law, which says that a front glide *j* assimilates to a following *i,* if preceded by a consonant within the same syllable".

(23) *Sievers' Law in Gothic*
ji → iː / C_1 —[19]
(Vennemann 1971)

This phonological rule is more interesting by what it excludes than by what it states. For instance, it does not apply to back glides and vowels.

(24) Gothic: *wu → uː / $^{\$}C_1$ —

This lack of parallelism in a phonological rule applying to suffixes becomes less mysterious if we compare the rule to a morpheme-structure condition of the language.

(25) Gothic: # Cwu, but *# C_1ji

Word-initial consonant-w-u sequences occur, while word-initial consonant-j-i sequences do not. "The rule is thus a very natural one in that it changes sequences arising morphologically which the speaker is not accustomed to from the word-initial position, namely $^{\$}C_1ji$ sequences, into something that is common in the language $^{\$}C_1i$: No such amendment is called for in the case of the velar glide. Note, incidentally that this explanation cannot even be expressed unless we use the concept of a syllable" (Vennemann 1972:11-12).

As the author himself notices, the rule is understandable and explained with reference to the morpheme-structure constraint and the word-initial well-formedness constraint. The question is then, why does the author need a syllable as a crucial unit? If we were to formalise this generalisation in terms of the syllable, we would certainly miss the point, and would obscure the motivation of the process rather than reveal it. The process in Gothic occurs at the juncture of a stem and an affix, and the sequence (containing a stem-final segment and affix-initial one) undergoes a change

[18] The Law of initials, as defined by Vennemann, says: "Medial syllable-initial clusters should be possible word-initial clusters" (Vennemann 1972:11).
[19] The symbol $ stands for a syllable boundary.

in case the created sequence (i.e. V-V) does not meet a morpheme-structure condition. Thus, what happens across morpheme boundary is in a causal relation to morpheme constraints. Such processes are quite common across languages. Consider the following epenthesis process in Georgian.

(26)

Stem	*Derived form*	
rje	me-rje-v-e	'milk'
t'χ'e	me-t'χ'e-v-e	'forest'
	(/me - e/ is an affix denoting a professional affiliation)	

In the examples in (26), epenthesis of the sonorant /v/ occurs in order to resolve vowel hiatus. The process cannot be explained without reference to the morpheme-structure conditions of Georgian, where stem patterns are enforced at the stem-affix boundary. The generalisation is that two adjacent vowels are disallowed in the monomorphemic context in Georgian.[20] The same constraint seems to operate across a stem and a suffix boundary. Remarkably, vowel hiatus between the prefix and a stem, or between a prefix and another prefix, is allowed.[21]

(27) a) prefix + prefix + root + suffix
c'a - a - k'itx – a 'made somebody read'
PREV - NEUT MARKER - V ROOT - PAST TENSE MARKER

b) prefix + root + suffix
sa - arak' - o 'fantastic'
(/sa - o/ - is an affix denoting abstract notions; /arak'-/ ROOT)

Thus, in the cases of both Gothic and Georgian, the formulation of a rule cannot be explanatory unless morpheme-structure conditions are mentioned.

Returning to the syllable, consider a process in Sanskrit discussed in Vennemann (1972).

> Sanskrit, as well as several related languages, has a rule by which a stop is doubled after an *r* and before a vowel: *mārggam*, *varggam*, *dīrggha*, *arkkaḥ*, *ārttaḥ*. This doubling seems to be in contradiction to the general rule of these languages that the first rather that the second of two consonants is doubled (cf. also the so-called West Germanic gemination): *addya*, *pretyia*, *puttrena*, *sappta-*, *akktubhiḥ*. The contemporary Indian grammarian Siddheshwar Varma gives an excellent explanation for the apparent conflict: "there is no doubt that in the above examples the first member of the group – viz., *r-* has not been

[20] There are, however, a few monomorphemic loan words with identical adjacent vowels in Georgian, e.g. /saati/ 'clock' (borrowed from Persian), /baasi/ 'conversation' (borrowed from Persian).
[21] Asymmetry between the behaviour of prefixes and suffixes has been attested in many languages (see e.g. Peperkamp 1997). In Chapter 3 I discuss some more phonological processes illustrating the prefix/suffix asymmetry in Georgian.

> doubled, and it may prima facie appear as if it was an exception to the rule of syllabification given above [the rule that, e.g., *putram* is *puttram* qua put $tram, T.V.]; but a closer reflection will show that even here the same tendency has worked – viz., that of keeping the first syllable close, so that *svarggam* was divided as *svarg/gam*. The first syllable here could not be closed with *r*, because there exists no word in Sanskrit that ends in *r* at the end of a sentence" (Vennemann 1972:13–14).

The generalisation in this case could be stated in terms of the syllable for the purpose of simplicity, but the generalisation would certainly miss the point, and would not be explanatory unless word well-formedness constraints are mentioned. Again, as in the case of Gothic, although the processes are argued to support the syllable as a unit, in fact, they refer to a word or a morpheme. Both cases are quite similar to the arguments for syllabification discussed in 2.1.2.1.

Often-cited evidence for the syllable is the process of vowel lengthening in Modern Icelandic, i.e. lengthening that occurs before medial consonants and consonant clusters. The process is illustrated by the examples in (28).

(28)	a)	hatur	[hɑ́ːtʰʏr̥]	'hated'
	b)	hattur	[hɑ́ʰtːʏr̥]	'hat'
	c)	ofsi	[ɔ́f.sɪ]	'violence'
	d)	titra	[tʰɪ́ːtʰra]	'shiver'

(Vennemann 1972)

A stressed vowel is long before a single intervocalic consonant, as in (28a), but short before clusters of two or more consonants, including geminates (see (28b) and (28c)). An exception to the rule is the example in (28d): a vowel lengthens before a cluster of /t/ and /r/. More generally speaking: a vowel lengthens exactly before the consonant clusters listed in (29).

(29) /p t k s/ + /r j v/

Vennemann (1972) suggests that the formulation of the rule could be what has traditionally been called "open syllable lengthening". The simple case of open syllable lengthening is demonstrated in (30).

(30) *Open syllable lengthening*

$$\begin{bmatrix} \text{V} \\ +\text{stress} \end{bmatrix} \rightarrow [+\text{long}] \,/\, \text{—}\ \text{CV}$$

This rule covers forms such as (28a), but not (28d). An additional rule such as (31) is needed to explain (28d).

(31) $\begin{bmatrix} \text{V} \\ +\text{stress} \end{bmatrix} \rightarrow$ [+long]/ — /p t k s/ + /r j v/ V

(30) and (31) are unified into a single rule in (32).

(32) $\begin{bmatrix} \text{V} \\ +\text{stress} \end{bmatrix} \rightarrow$ [+long] / — C_1 (C_2)V
Condition: C_2 = /r/ or /j/ or /v/ and if C_2
then C_1 = /p/ or /t/ or /k/ or /s/

Vennemann suggests that all his formulations are inadequate, because in all descriptions the real generalisation is missed. The real generalisation is that vowel lengthening takes place in an open syllable. He argues for the formulation given in (33).

(33) *Open syllable lengthening*

$\begin{bmatrix} \text{V} \\ +\text{stress} \end{bmatrix} \rightarrow$ [+long] / — $

However, the question remains why the clusters in (28) behave in this way in Icelandic. Vennemann proposes an informal and tentative answer: “these clusters are composed of one of the four strongest consonants of the language, followed by one of the three weakest” (Vennemann 1972:6).

The concept of a strength hierarchy is a traditional one. Unlike Foley (1970) (who bases his strength hierarchy on sound changes) and Sigurd (1955) (who bases the strength scale on clustering behaviour), Vennemann proposes a strength scale based on synchronic phonological rules, including syllabification rules. Consider some of the arguments as proposed by Vennemann.

In Icelandic the strong consonants /p/, /t/, /k/ and /s/ manifest their strength in the following ways:

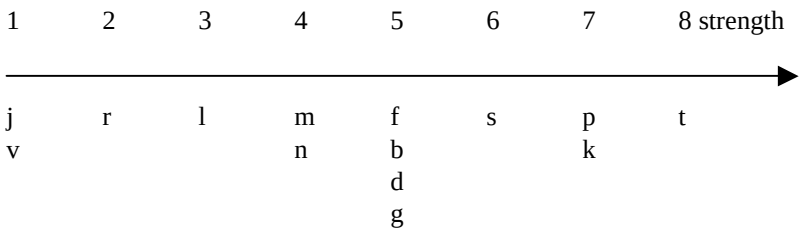
(34) a) /p t k/ are either aspirated or preaspirated, which no other consonants are.
b) /s/ is never voiced in voiced environment, while all other spirants are.

The relative weakness of /r/ as opposed to /l/, /m/ and /n/ is apparent from phonological processes such as:

(35) a) /r/ is always devoiced when following /p t k/; with other resonants devoicing occurs occasionally.
b) /r/ is always devoiced when following /s/; the other resonants never are.

Vennemann (1972) proposes a strength scale for the consonants of Icelandic.

(36) Relative strength of consonants in Icelandic (tentative) (Vennemann 1972:6).



As Vennemann points out, languages have a certain amount of freedom in the definition of their strength hierarchies. "Thus, while in Latin *l* is on the weak side together with *r*, and *s* is on the weak side as opposed to the stops, *s* is on the strong side together with the voiceless stops in Icelandic, and *l* likewise is there on the strong side, as opposed to *r*" (Vennemann 1972:11).

Thus, although the formulation of open syllable lengthening in (33) seems simple and transparent, in fact, the point is missed if the 'strength' of segments, an apparently language-specific matter, is not taken into account. One might then ask: what is the advantage of (33) over (32) and (31)? After all, whether explicitly (as in the formulation of the rule in (31) or (32)) or implicitly (as in (33)), we need to refer to the concrete segments involved in the process. The same could also apply to the argument concerning weak vs. strong clusters (the division was originally suggested by Chomsky & Halle 1968).

Vennemann seems unconvinced by his arguments since he says: "I will advocate here the incorporation of syllable boundaries and syllables in phonological description. I will not say, however, that the incorporation of these concepts into the theory of grammar is 'necessary'"(Vennemann 1972:2).

Another often-cited example is that of German final devoicing. As argued by many, the process can be stated only with reference to a syllable, not to a word, morpheme or stress (for a different opinion see Kasevič 1986, Butskhrikidze 1994, 1998a). Consider the examples in (37).

(37)	a)	Liebling	/liːb.lɪŋ/	[líːplɪ̀ŋ]	'darling'
	b)	lieble	/liː.blə/	[líːblə]	'I flirt'

The devoicing of [b] takes place in (37a), but not in (37b). Vennemann says that behind the difference lies a different syllabification pattern of the words: *Liebling* is syllabified as /liːb.lɪŋ/ while *lieble* /lieble/ is syllabified as /lie.ble/. As Vennemann notes: "this difference in syllabification is undoubtedly related to stress, since *Liebling* has a secondary stress on the suffix, *lieble* does not" (Vennemann 1972:8). Note that the devoicing is not explained by direct reference to a syllable as such, but via the stress pattern. The devoicing is the rule primarily applying in word-final position in German, as shown in the examples in (38).

(38) gelb [gɛlp] ‘yellow’
tag [tak] ‘day’
rad [rat] ‘bicycle’

Word-final devoicing is attested in many languages, e.g. in Dutch, Polish and Russian.

Insofar as I am arguing against the role of the syllable in underived words, good evidence for this position would be a language which has devoicing only in syllable-final position of underived words. Such evidence would demonstrate that the devoicing is exclusively a syllable-related process. I am not aware of the existence of such a language, or of a study reporting such a case.[22] However, languages where word-final devoicing occurs only in derived forms do exist, e.g. Georgian (for more about the process see Chapter 3).

Thus, I conclude that the phonological rules discussed so far do not necessarily require reference to the concept of syllable.

2.1.2.5. Accent assignment

The most convincing arguments for the syllable as a domain concern accent assignment (Liberman & Prince 1977, Selkirk 1980, Hayes 1981). There are languages in which the weight of the syllable determines accent placement (e.g. Khalkha Mongolian, Huasteco, Cuna). Accent assignment in such languages is sensitive to the distinction between heavy and light syllables, where ‘heaviness’ is defined on a language-particular basis, e.g. some count CVV as well as CVC syllables as heavy, while others count only CVV syllables as heavy.

There are also languages where accent placement is determined not by the weight of a syllable, but by the morphological structure of a word, e.g. the stem domain. In such languages, accent always falls on the stem and never on a grammatical affix, e.g. Yupik languages (Jacobson 1985), Georgian (Butskhrikidze 1996). For more about stem stress languages and the Stem Stress Principle see Butskhrikidze (1996), van de Vijver (1998) and Revithiadou (1999).

In yet other languages, accent always falls either on the first or on the last syllable, regardless of the syllable weight or of the morphological composition of a word, e.g. word-initial accent is attested in Czech, Icelandic and Finnish. Word-final accent exists in Persian and Turkish.

Since there is only one primary accent in an underived word, one cannot say whether the syllable or stem can serve as a domain which is independent from the

[22] However, Jeroen van de Weijer has drawn my attention to one example in Dutch where devoicing takes place syllable-finally in an ‘underived’ word: *cadmium* ‘cadmium’ is pronounced as [ˈkat.miɔm]. Even though there is no independent morpheme *cadm*, the form *cadmium* maybe considered to be morphologically complex; it contains a fairly productive suffix *-ium*, attested in forms such as *calcium*, *chromium*, *barium*, *atrium*, etc. Thus, the devoicing in *cadmium* is not a strong piece of evidence against claims made above.

word. Both the syllable and the stem are sub-constituents of the word, and accent should primarily be characterised as a property of the word as a whole.[23]

2.1.2.6. Conclusions

In this section I have discussed whether the syllable can function as an independent domain from the word-level, and hence, for word-level phonotactics. I have looked at the issue from the following perspectives.

- Syllabification
- Phonotactics
- Word-edge phenomena
- Phonological processes
- Accent assignment

None of these could substantiate the crucial role of the syllable for the word level. Note that since the present study is restricted to underived words, I have not investigated the role of the syllable for derived words or for syntactic structures (e.g. processes like *raddoppiamento sintattico* in Italian or *liaison* in French). For more about these issues see Nespor & Vogel (1982), Peperkamp (1997) and Ewen & van der Hulst (2001).

I return now to the generalisation for Georgian in (6). For ease of exposition I repeat the generalisation here.

(39) The vowel /u/ is commonly found in the final position of monosyllabic words; however, its occurrence in the final position of polysyllabic words is very restricted.

Earlier, I used the generalisation in (39) to argue for smaller units of a word serving as a domain for phonotactic generalisations. Note that even though the generalisation refers to the number of syllables in a word, it does not support the claim that the syllable is an independent domain. In both parts of the generalisation we are still dealing with the word (i.e. monosyllabic word and polysyllabic word); in both cases the syllable is used only as shorthand notation.

To conclude, the different phonological criteria discussed in this section have shown that the syllable does not comply with the requirement formulated in (10b), and consequently, cannot serve as the domain of phonotactic generalisations.

In the next section, I turn to the stem and examine whether it can serve as an independent domain for phonotactic and other phonological generalisations.

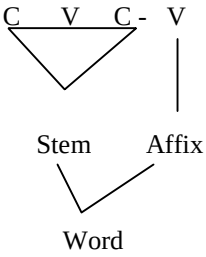
[23] Note the existence of an implicational relation between the heavy syllable, stem and word. Commonly, heavy syllables are found in lexical rather than in grammatical morphemes. Thus, the heavy syllable constitutes a lexical morpheme, which is an obligatory constituent of a word.

2.1.3. The stem

In this section I propose that the stem can be defined as a phonological unit, and eventually integrated into the phonological hierarchy.

Since the stem constitutes a meaning-related aspect of the word, viewing word phonotactics in terms of the stem implies considering the lower part of the representation in (9), repeated in (40).

(40)



As seen from the schema in (40), the stem could potentially be an independent unit of the word. This is primarily due to the fact that the word consists of an optional affix, as well as the stem.

Let us examine the independent status of the stem by looking at different processes, similar to the ones which were looked at earlier in order to reveal the status of the syllable.

(41) a) Phonotactic restrictions
b) Phonological processes
c) Accent assignment

Evidence for each of these is discussed in the following subsections.

2.1.3.1. Phonotactic restrictions

It is no coincidence that the person who invented the term ‘morpheme’, the Polish linguist Baudouin de Courtenay, was also the first to recognise the morpheme as a domain of phonotactic restrictions (Baudouin de Courtenay 1963). A large number of linguists have studied the phonotactics of a language by considering the stem/root as the most appropriate domain. Such studies include Semitic (Greenberg 1950), Proto-Austronesian (Chrétien 1965), Indo-European (Magnusson 1967), Polynesian (Krupa 1971), Avarian (North-Caucasian) (Mdivani 1983), Uralic (Bakró-Nagy 1992) and Kartvelian (Melikishvili 1997). The methodology also seems to underlie *The sound pattern of English* (Chomsky & Halle 1968), the pioneer work in Generative phonology, which assumes that the phonotactics of a language is accounted for by two mechanisms: morpheme-structure conditions (MSCs) and phonological rules. Within the generative framework, morpheme-structure conditions apply to the underlying, lexical representations of morphemes, and phonological rules derive the

surface forms from the underlying forms. Together, they define the possible combinations of sounds of a language (Postal 1968). Since two levels have been mentioned, I would like to emphasise that the phonological stem domain which I argue for in this thesis is a surface matter and that co-occurrence constraints holding within this domain are output constraints. The most recent work arguing for the importance of the stem domain in phonological descriptions, as well as in phonological theory, includes Inkelas (1990), Borowsky (1993), the framework of Optimality Theory (with the introduction of alignment constraints; McCarthy & Prince 1993), Downing (1999) and Booij (2000). The issue of incorporating the stem domain in phonological theory is largely dependent on how one sees the morphology-phonology interface within the language faculty. I return to this issue at the end of this chapter.

Firstly, let us present some empirical data illustrating the stem as the domain of phonotactic generalisations. Shibatani (1973) has pointed out that distributional constraints on morphemes often differ from constraints characterising the word. For instance, a morpheme in Japanese may end in any of a number of consonants, but a word must end either with a vowel or with a moraic nasal. Recall that a similar constraint applies to Georgian and Italian (see section 2.1.1).

Some studies report that special features hold only at the lexical morpheme level. For example, in Chaoyang, a dialect of Chinese, the features [constricted glottis] and [nasal] are specified only at the level of the lexical morpheme (Yip 1994). Similarly in Malayalam, a single dental nasal followed by a vowel can only occur morpheme-initially; in this environment, an alveolar nasal cannot occur (Mohanan 1995). In languages with vowel harmony certain vocalic features are inherent only in lexical morphemes. For instance, in Turkic languages the opposition back vs. front (partly also rounded vs. unrounded) and in Tungus and other languages in the Far East the opposition high vs. low (Jakobson 1949) only play a role in lexical morphemes.

The generalisations concerning the occurrence of special features or segments only within the lexical morpheme domain are part of a more general observation concerning the asymmetry between the phonotactics of lexical vs. grammatical morphemes. Consider the case of Dutch: lexical morphemes cannot begin with a schwa-headed syllable,[24] however, grammatical affixes can have schwa as their only vowel. The data are taken from Booij (2000).

[24] There are few exceptions, but such words have a polymorphemic origin: the schwa-syllable is historically a prefix, *be-* or *ge-*. There are some French loans in which the unstressed mid vowel has been reduced to a schwa, e.g. *beton* 'concrete', *debacle* 'failure', *gelei* 'jelly'.

(42)	gebelazer	g[ə]-b[ə]-laz[ə]r	‘cheating’
	geversier	g[ə]-v[ə]r-sier	‘decorating’
	gelukkiger	g[ə]-lukk-[ə]g-[ə]r	‘happier’
	verrukkelijk	v[ə]r-rukk-[ə]l[ə]k	‘delightful’
	wandelen	wand[ə]l-[ə]n	‘to walk’
	lekkerder	lekk[ə]r-d[ə]r	‘nicer’

The same type of constraint on the occurrence of schwa in lexical vs. grammatical morphemes is attested in German (see Hall 1999).

The cross-linguistic observation that non-lexical morphemes, i.e. affixes, obey specific phonotactic constraints which are not valid for lexical morphemes has been made by Jakobson (1949), Nida (1949), Hyman (1975), Booij (1985) and Zubkova (1988), among others. For instance, the native prefixes of Dutch consist of at most one syllable, and the native suffixes of at most two, of which only one contains a full vowel. Lexical morphemes do not conform to these restrictions (Booij 1977:22–23). The asymmetry long stem vs. short grammatical morpheme is a well-acknowledged cross-linguistic fact. The existence of such a correlation is attested in a number of languages. Dirks (1953) reports the existence of such a correlation for Campa Arawak, Russell (1999) for Cree and Hall (1999) for German.

Another type of asymmetrical pattern concerns the difference in the use of phonemic inventories. Consider the following observations, for instance: of the twenty-three consonants in spoken Czech only eight phonemes are used in inflectional suffixes. Three of these appear in nominal endings and six in verbal ones; /m/ is the only consonant that occurs in both of these cases. “Only an insignificant percentage of English phonemes participate in inflexional suffixes: there occur only four consonantal phonemes: *z, d, n, and ŋ.* Both the vowels of all these suffixes and the unvoiced variants of the suffixes *-z* and *-d* are automatically conditioned by he preceding phoneme and have no distinctive value” (Jakobson 1949:108). Of the twenty-eight consonants in Modern Georgian only eight phonemes are used in inflectional morphemes. Interestingly, these types of generalisations show that the phonemes that appear in grammatical affixes are a subset of the phonemes that occur in lexical morphemes. The consonantal inventory of grammatical morphemes presents the least marked patterns. For instance, in grammatical morphemes, the commonly attested consonants are /b-p d-t g-k s m n l-r/. In the Finno-Baltic languages the sounds which commonly occur in grammatical affixes include /t n k l/, in German: /m n r t s/ and in Arabic: /n l m s t n/ (Zubkova (1990). Phonemes with a complex structure, e.g. affricates, or sounds with a secondary articulation are not usually found within a grammatical morpheme. For instance, labialised consonants appear only in lexical morphemes in Archi (Zubkova 1990), and, similarly, pharyngealised consonants appear only in lexical morphemes in Arabic (Mel’nikov 1966). The restriction does not apply to lexical morphemes, which generally exploit the entire phoneme inventory of a language. The asymmetry in the phonemic constituency of lexical and grammatical morphemes correlates directly with the fact that consonant

clusters are generally found within lexical morphemes rather than in grammatical ones (Butskhrikidze 1998a). I will return to this correlation later.

Moreover, co-occurrence restrictions, e.g. phonological principles such as the SSP, the OCP and others, operate in the stem domain. These principles and data are discussed in section 2.3.

2.1.3.2. Phonological processes

As was the case with phonotactic restrictions, the aim of this section is to illustrate that the stem functions as a domain which is independent from the word. Consider the case of Cree (an Algonquian language). The data are taken from Russell (1999). Cree is a polysynthetic language; for example, a general 'template' of the Cree verbal complex is a prefix (a person marker) followed by zero or more preverbs, the stem and a number of inflectional suffixes. According to Russell (1999), based on patterns of phonotactics and accent assignment, preverbs constitute phonological words of their own. The process relevant to the present discussion is word-final devoicing. Word-final vowels in Cree optionally devoice towards the end of the vowel. The partial devoicing is indicated by a word-final *h*.

(43)	anima ~ animah	[anɪma] ~ [anɪmaḁ]	'that one'
	ahpô ~ ahpôh	[aḁpoː] ~ [aḁpoːo̥]	'even'
	pîhtikwê ~ pîhtikwêh	[piːi̥tɪgweː] ~ [piːi̥tɪgweːe̥]	'enter!'

The final vowels of preverbs are also optionally devoiced towards the end.

(44)	kî ~ kîh	[kiː] ~ [kiːi̥]	'past'
	kita ~ kitah	[kɪta] ~ [kɪtaḁ]	'in order to'

This contrasts with those environments where optional devoicing does not occur, such as the boundary between the stem and a suffix (see (45)) or between the two halves of a compound (see (46)).

(45)	ê-nipâ-c-ik	[eː nɪpaːtʃɪk]	'that they sleep'
	COMP-sleep-3-3PL	*[eː nɪpaːḁtʃɪk]	
(46)	kinosêwi- sîpîhk	[kɪnʊseː.wɪ.siːpiːxk]	'at Norway house'
	fish-river (LOC)	*[kɪnʊseː.wɪi̥.siːpiːxk]	

As shown in these examples, final vowel devoicing is characteristic exclusively of the phonological word domain, not of the stem. Stem-final vowels do not undergo devoicing.

Another process sensitive to the stem boundary in Cree involves external sandhi. Preverbs, as in the previous cases, behave identically to independent words with respect to sandhi phenomena. When two vowels are adjacent across a word bound-

ary, they may undergo optional coalescence or deletion of the first with compensatory lengthening of the second.

(47)	êkwani	anima	>	êkwanânima
	and	that:one		

This sandhi process does not apply across a boundary between a stem and a suffix. In this environment, vowel hiatus is resolved in other ways, depending on the identity of the morphemes involved, e.g. by epenthesis of /y/ (Russell 1999:210).

The generalisations in Cree, which distinguish two environments, the full word boundary and the word-internal (or stem) boundary, are well attested in other languages, and are formalised in various ways, depending on the theoretical framework, e.g. in terms of boundary symbols, by distinguishing between the full word (external) ## vs. # stem boundary (see Hyman 1975). In early Generative frameworks, the difference was translated into different morphologically sensitive bracketings of the string. In stratal theories such as Lexical Phonology, Level 1 and Level 2 domains are distinguished to account for different phonological outputs. Although different formalisations and strategies are employed in these theories, the empirical fact nevertheless remains invariant. Processes in many languages require explicit reference to the stem boundary, and thus to the stem domain (for instance, palatalisation in Korean (Hyman 1975) and gemination in Malayalam (Mohanan 1995)).

Another case where the stem functions as an independent domain is vowel harmony. Studies on vowel harmony phenomena repeatedly report that a majority of languages having vowel harmony are of the stem-controlled type (van der Hulst & van de Weijer 1995). In languages such as Tangale, a Chadic language, or Finnish the stem vowels are invariant, while the affixal vowel varies, depending on the stem vowel. In Tangale, when a stem has a closed vowel, the vowel of the affix is also closed, and when a stem has an open vowel, the vowel of the affix is also open. The examples are taken from van der Hulst & van de Weijer (1995).

(48)	seb-u	‘look’ (IMP)
	kɛn-ʊ	‘enter’ (IMP)
	tug-o	‘pounding’
	wʊd-ɔ	‘farming’

Finnish vowel harmony is based on the front-back opposition. As expected, affixes show a front or back alternant depending on the quality of the stem vowel.

(49)	tyhmä-stä	‘stupid’ (ILL)
	tuhma-sta	‘naughty’ (ILL)

In line with this type of generalisation, allomorphic alternations of suffixes triggered by the phonotactic patterns of a stem are attested in many languages, e.g. umlaut in

German and Icelandic, and stem-dependent allomorphic alternations in Dutch and Georgian. As an illustration of Icelandic umlaut, consider the examples in (50).

(50)

kalla	köllum	'call, 1 SG vs. 1 PL'	
saga	sögu	'story, NOM vs. ACC'	(Árnason 1985)

As an illustration of the stem-dependent process, consider the following case from Dutch. The first segment of the diminutive suffix /-tjə/ is homorganic with the preceding final nasal segment of the stem, e.g. *riempje* [rimpjə] 'little belt' derives from /rim-tjə/ and *koninkje* [konɪŋkjə] 'little king' derives from /konɪŋ-tjə/ (Booij 1985). Thus, the allomorphic pattern of the grammatical morpheme is stem-dependent.

Consider now the case of Georgian. Stems in Georgian can end either in a vowel or in a consonant, while words end only in a vowel. Nominal word formation is stem-dependent; more specifically, the affixes attached to the stem alternate depending on whether the stem ends in a vowel or in a consonant. For example, allomorphs for the ergative case are /-m/ and /-ma/, respectively.

(51) a) *Vowel-final stems*

NOM	*ERG*	
xe	xe-m	'tree'
sok'o	sok'o-m	'mushroom'
deda	deda-m	'mother'

b) *Consonant-final stems*

NOM	*ERG*	
k'ac-i	k'ac-ma	'man'
saxl-i	saxl-ma	'house'
c'ign-i	c'ign-ma	'book'

As shown in the examples in (51), the allomorph /-m/ is chosen when the stem ends in a vowel (see (51a)), while the allomorph /-ma/ is chosen when the stem ends in a consonant (see (51b)).

This type of word formation, which is characteristic of Georgian, is called stem-inflection in Bloomfieldian terminology. Other languages that belong to this type are Latin and German, contrasting with word-inflection languages, e.g. English (Bloomfield 1933). The difference was later formulated as word-based vs. stem-based morphology (Wurzel 1989). A language in which realisation rules are usually defined on stems that occur as free phonological words will have word-based morphology and those whose realisation rules are defined on stems that do not occur as free forms have stem-based morphology. As one may notice, in both formulations we are dealing with a single unit: the stem. This is the reason why Aronoff (1994) finds this kind of distinction between the two types inadequate, and instead proposes to account for both phenomena with a unitary term: *lexeme*. The term *stem* in his approach is merely a sound-form aspect of the lexeme. Discussing the details of

these approaches is beyond the scope of this thesis, and I will not explore them in detail any further. The important point is that the stem domain is different from the word domain, and recognising the stem domain in language modelling is necessary not only for accounting for sub-word phenomena, but for the formalisation of word-formation processes as well. The function of the stem in the former, not in the latter, environment is the focus of this thesis.

The aim of the section was to demonstrate that the stem functions as an independent domain. Various phonological processes have shown that it certainly does.

2.1.3.3. Accent assignment

Chomsky & Halle (1968) was the first work to offer the formalisation of accent assignment by taking into account the morphological constituency of a word. In morphological systems (according to Hayes' 1995 terminology) such as that of English, the main stress is on the lexical morpheme, while affixes are stressless. Thus, "… the fact that *un # bound # ed # ness* has antepenultimate stress has nothing to do with any rhythmic principles, but merely reflects the fact that the stem syllable is in antepenultimate position in this word" (Hayes 1995:32). Two types of morphologically sensitive language can be distinguished: (i) languages in which the stress always falls within the stem domain regardless of the type of affix attached to it, and (ii) languages in which the stress can shift from the stem to the affix, depending on the type of affix.

Georgian and Norton Sound belong to the first group of languages. The same holds true for some Bantu languages, where high tone is restricted exclusively to the stem domain.

Consider, for instance, the case in Norton Sound, a dialect of the Central Alaskan Yupik language. The data are taken from van de Vijver (1998).

(52)	*Stem*	*Stem + suffix*	
	átəq	átxani	'in his name'
	kúvə	kúvvuq	'it spills'
		kúvluni	'(it) spilling'

Consider now the case of Georgian. As in Norton Sound, the accent is stem-bound.

(53)	*Stem*		*Stem + suffix*	
	déda	'mother'	déd-eb-i	'mother PL'
	táro	'shelf'	táro-eb-it	'with shelves'
	sók'o	'mushroom'	sók'o-ian-i	'made of mushroom'

Modern Greek, English and Russian belong to the second group of languages.

Consider the case of Greek. The data are taken from Revithiadou (1999).

(54) *Paradigm of nouns*

a)	klívan-os (NOM SG)	'furnace'
	klívan-u (GEN SG)	
	klívan-i (NOM PL)	
	klívan-on (GEN PL)	
	klívan-us (ACC PL)	
b)	ánθrop-os (NOM SG)	'man'
	anθróp-u (GEN SG)	
	ánθrop-i (NOM PL)	
	anθróp-on (GEN PL)	
	anθróp-us (ACC PL)	
c)	θálas-a (NOM SG)	'sea'
	θálas-as (GEN SG)	
	θálas-es (NOM PL)	
	θalas-ón (GEN PL)	

In the examples in (54a), the accent is stable; it always falls on the same syllable of the stem. In the examples in (54b), the accent shifts from antepenultimate to the penultimate syllable, but remains within the stem domain, while in the examples in (54c), accent shifts from the stem to the GEN PL marker /-on/.

The data I have discussed on accent assignment illustrate that in many languages accent location is morphologically determined. It is predominantly stem-bound; however, there are also languages in which affixes cause a shift of the accent from the lexical to the grammatical part of a derived word. In either case, the formalisation of these facts necessarily calls for the incorporation of morphological constituents, such as the stem and the affix, into the phonological theory. The prosody-morphology interface, in term of a theory of *head dominance* and principle of *prosodic compositionality*, is fully articulated in Revithiadou (1999).

2.1.3.4. Conclusions

Various phenomena have been examined in order to establish the independent status of the stem domain.

a) Phonotactic restrictions
b) Phonological processes
c) Accent assignment

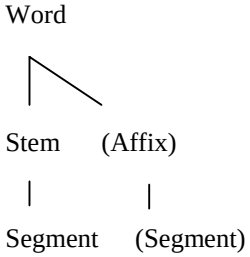
The examination of each of these aspects has shown that the stem functions as an independent domain. Since both requirements for the sub-word domain have been met (recall the requirements formulated in (10)), the introduction of the stem domain in phonological theory can be justified. Recognising the stem domain means recognising, in parallel fashion, the affix domain. However since, as argued above, affixes show patterns which are sub-classes of the stem patterns, and since this thesis does not deal with derived words, I concentrate on the stem domain. In addition, the focus

on the stem is justified for the reason that the thesis primarily focuses on consonantal complexity, a feature characteristic of the stem domain.

2.1.4. General conclusions

To conclude, the units relevant for word-level phonotactics are: the word, the stem (and an optional affix) and the segment. The internal hierarchical organisation of the word is depicted in (55).

(55)



The unidimensional hierarchy in the schema in (55) incorporates units of a different nature: the segment is merely a formal part of the word, while both the stem and the word consist of indivisible formal and meaning components. Thus the hierarchy portrays the essence of the matter inadequately.

Towards the end of this chapter (see (96)), I suggest a revised version of the hierarchy, a three-dimensional representation of the word-internal structure incorporating formal, meaning and structural components of the word-domain.

As far as the importance of the morphological constituency of the word has been recognised, a short excursion into morphological typology is appropriate at this point. The next section deals with the morphological classification of languages, which serves as background for the discussion of the principles later in this chapter.

2.2. Morphological typology

2.2.0. Introduction

In Chapter 1, I argued that the asymmetry between the lexicon and the grammar plays an important role in the study of phonotactics. Besides general characteristics (e.g. length and phonological constituency) associated exclusively with lexical and grammatical morphemes, it is essential to explore how the lexicon and grammar interact within a specific language. To put it another way, it is important to know where the division line between the lexicon and grammar is in a given language. Previous studies on morphological typology have shown that the issue bears on the type of morphology a language possesses. In the following section I will demonstrate how the patterns of the interaction can affect the phonotactic organisation of a word. First, a brief introduction to the history of the study of morphological typol-

ogy is in order. It is followed by a revised version of the morphological classification, which is proposed in order to demonstrate more vividly the correlation between the morphological type and the phonotactics.

2.2.1. Previous studies on morphological typology

The foundation of morphological typology was laid at the beginning of the twentieth century. At this early stage, two factors were used to characterise languages.

(56) a) The extent to which linguistic concepts are expressed by morphology (i.e. word-internal modifications), rather than by the use of separate words.
b) The morphological technique employed.

These two factors were typically conflated into a simple tripartite or quadripartite classification of languages. August Wilhelm von Schlegel (1918), developing the work of his brother, Friedrich, proposed a tripartite division into the following types.

(57) a) Languages without any grammatical structure.
b) Languages which employ affixes.
c) Languages with inflection.[25]

Languages like Chinese are of the first type, in which roots are generally not modified by affixation or internal change and words therefore appear to lack any 'grammatical' structure.

As an example of the second type, Schlegel cites some of the American-Indian languages and Basque. Basque, for instance, has a complex system of cases, which are expressed as suffixes. Despite the apparent complexity of the paradigm, the endings represent individual cases and are mostly invariant and identifiable. This seems to be the group of the classical agglutinative languages, e.g. Turkish and Finnish.

To the third group belong languages like Old Indo-European and Modern Indo-European. According to Schlegel, characteristic of this type is that the roots themselves can be modified in order to express grammatical categories, with or without the concomitant presence of affixes (Collinge 1990).

The tripartite classificatory typology of Schlegel was expanded into a quadripartite typology by introducing 'incorporating' languages as a separate language group (Humboldt 1922).

[25] Schlegel further divided the inflectional languages into analytic and synthetic languages.

(58) a) Isolating languages (e.g. Chinese).
b) Agglutinative languages (e.g. Basque).
c) Flectional languages (e.g. Greek and Latin).
d) Incorporating languages (e.g. American-Indian languages).

Many attempts have been made to improve the classical nineteenth century morphological classifications. In the twentieth century the most notable works were those of Sapir (1921), Skalička (1979) and Sgall (1986).

Later, it was commonly acknowledged by typologists that there is no language that can be characterised as being solely one of the four types in (58), i.e. there is no 'pure' language. For instance, even Chinese is not entirely an isolating language; it possesses a number of grammatical suffixes, for example the aspectual suffixes /-le/ (perfective) and /-zhe/ (durative), which are attached agglutinatively to verbal roots (Li & Thompson 1987).

The recognition of the relative nature of the classification led many to apply a different methodological technique, namely, that of the quantitative approach. The method was first applied by Greenberg (1960) and followed later by Cowgill (1963), Mejlax (1973) and Kasevič & Jakontov (1982). In quantitative studies, special indexes (e.g. an index of agglutination, inflection, etc.) are calculated for each language. Languages are compared in terms of these indexes.

The morphological type is what Sapir (1921) calls the 'structural genius' of a language, i.e. the basic plan of a language. Obviously, there are certain patterns in which languages differ, and patterns which are common across languages. The issue is where the differences lie and which terms should be used to express those differences adequately. There are many different stances one could take in approaching the issue. One way is proposed in the following section. Since it is impossible to set up a limited number of types that would do full justice to the peculiarities of all languages spoken in the world, in what follows I describe mere tendencies.

2.2.2. The morphological classification revised

Sapir (1921) used two parameters in his morphological classification: (i) quantitative and (ii) qualitative, each characterising a distinct word construction. The quantitative parameter (i.e. degree of synthesis) refers to number of morphemes in a word. Three types of word constructions are distinguished:

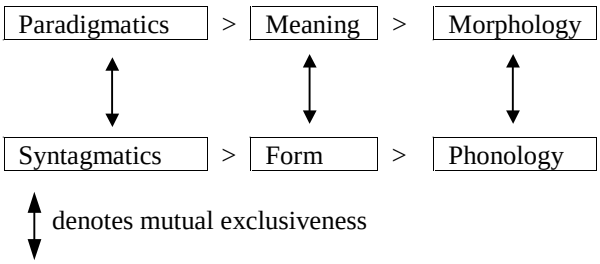
(59) a) Analytic, in which the lexical morpheme coincides with a word.
b) Synthetic, in which a word consists of a stem and an obligatory affix (either prefix or suffix).
c) Polysynthetic, in which a word may consist of several affixes.

The qualitative parameter (i.e. degree of fusion) refers to the ways morphemes are combined in a word. Four types of constructions are distinguished:

(60) a) Isolating, in which lexical morphemes do not combine with grammatical morphemes, are formally independent and thus easily identifiable.
b) Agglutinative, in which lexical morphemes combine with grammatical morphemes, but remain autonomous and are easily separable.
c) Symbolic.
d) Fusional.
In the last two types (symbolic and fusional) lexical and grammatical morphemes are combined in such a way, that it is sometimes difficult to separate them from each other. Additionally, grammatical morphemes of these constructions are characterised by polysemy, which makes their identification even more difficult.

The quantitative and qualitative parameters used for the classification of word constructions can be equated with the two possible dimensions of the study of language structure: syntagmatics and paradigmatics, respectively. Syntagmatics is understood as relations among elements in a sequence (i.e. horizontal relations), while paradigmatics is understood as relations among elements displaying some kind of kinship in some kind of parameter (i.e. vertical relations). The former is primarily characteristic of formal aspect of language, while the latter is primarily characteristic of meaning-related aspect of language. The approach, which would treat syntagmatics and paradigmatics as independent domains of study, would yield a linguistic theory treating phonology and morphology as mutually exclusive domains, as presented in (61).

(61)



As seen from the schema, linguistic theories which view paradigmatics and syntagmatics as mutually exclusive domains of the study also consider meaning and form, as well as morphology and phonology, as separate domains of investigation. As argued in Chapter 1, the thesis advocates that syntagmatics and paradigmatics are interdependent and cannot be studied independent from each other. So are form and meaning and phonology and morphology.

Sapir (1921), recognising the interdependence between syntagmatics and paradigmatics, had conflated the qualitative and quantitative aspects and had offered the following classification of languages:

(62) *Language classification* (Sapir 1921)

a) Pure-relational languages: Simple
Complex

b) Mixed-relational languages: Simple
Complex

Simple pure-relational languages are those that keep the syntactic relations pure and that do not posses the power to modify the significance of their radical elements by means of affixes or internal change. They can also be referred to as pure-relational or derivational languages.

Complex pure-relational languages are those that keep the syntactic relations pure and that also possess the power to modify the significance of their radical elements by means of affixes or internal change. They can also be called pure-relational deriving languages.

Simple mixed-relational languages are those in which the syntactic relations are expressed in necessary connection with concepts that are not utterly devoid of concrete significance but that do not, apart from such mixture, possess the power to modify the significance of their radical elements by means of affixes or internal change.

Complex mixed-relational languages are those in which the syntactic relations are expressed in mixed form, as in simple mixed-relational languages, and that also possess the power to modify the significance of their radical elements by means of affixes or internal change.

For instance, consider the following language classification proposed by Sapir (1921):

(63) *Language classification* (Sapir 1921)

Fundamental type	Technique	Synthesis	Example
a) Simple pure-relational	Isolating	Analytic	Chinese, Annamite
	Isolating (weakly agglutinative)	Analytic	Ewe (Guinea Cost)
	Agglutinative	Analytic	Modern Tibetan
b) Complex pure-relational	Agglutinative-isolating	Analytic	Polynesian
	Agglutinative	Synthetic	Turkish
	Fusional- agglutinative	Synthetic (mildly)	Classical Tibetan
c) Simple mixed-relational	Agglutinative	Synthetic	Bantu
	Fusional	Analytic (mildly synthetic)	French
d) Complex mixed-relational	Fusional- agglutinative	Polysynthetic (mildly)	Chinook (lower Columbia River)
	Fusional	Analytic	English

(Sapir 1921:150–151)

In line with Sapir's classification, in which paradigmatic and syntagmatic aspects of language intersect, I offer a classification with a primary focus on the relation between lexicon and grammar within a language. The basic idea, which also underlies Sapir's classification, is that all languages need to express radical concepts and relational ideas and what languages choose to adopt is a question of tendency. The way morphemes combine in a word directly bears on the problem of their formal independence. Formal independence of a word directly relates to the phonotactic organisation of a word. Therefore, formal characteristics of the sub-constituents of a word are the main parameter of the classification. Three main groups of languages can be distinguished:

(64) a) Lexical
b) Lexico-grammatical
c) Grammatical

Lexical languages are characterised by the formal coincidence of a word with a lexical morpheme. Grammatical morphemes similar to lexical morphemes occur as independent words. To this group belong traditional isolating languages, e.g. Chinese and Vietnamese.

Lexico-grammatical languages are also characterised by the formal coincidence of a word with a lexical morpheme. Unlike Lexical languages, grammatical morphemes of this language type sometimes do and sometimes do not appear as independent words. In either case they are easily identifiable: having a one-to-one mapping between their forms and meanings. Thus, such languages have independent stem ands dependent affixes. To this group belong traditional agglutinative languages, e.g. Turkish and Finnish.

Grammatical languages are characterised by the formal non-coincidence of a word with a lexical morpheme. Grammatical morphemes rarely occur independently. Identification of both lexical and grammatical morphemes is difficult. Thus, such languages have dependent stems and dependent affixes. To this group belong traditional inflectional and incorporating languages, e.g. Latin and Russian.

Word constructions in these languages are as follows:

(65) a) Lexical
#lexical morpheme# = #word#
#grammatical morpheme# = #word#

b) Lexico-grammatical
#lexical morpheme# = #word#
#grammatical morpheme# ≠ #word#

c) Grammatical
#lexical morpheme# ≠ #word#
#grammatical morpheme# ≠ #word#

The formal independence of lexical and grammatical morphemes in these three types of languages is summarised in (66).

(66)

Type of morpheme / Type of language	Lexical morphemes	Grammatical morphemes
Lexical	Independent	Independent
Lexico-grammatical	Independent	Dependent
Grammatical	Dependent	Dependent

Thus, as shown in the table, the formal independence of both lexical and grammatical notions is found in Lexical languages, while the formal dependence of both lexical and grammatical notions is found in Grammatical languages.

What does this say about phonotactics? Zubkova (1990) has proposed that phonotactic patterns, specifically the consonantal constituency of CVC words, vary depending on language type. She has studied consonant co-occurrence restrictions in C_1VC_2 words in three languages: Vietnamese (Lexical), Turkish (Lexico-grammatical) and Russian (Grammatical). The following phonotactic patterns have been observed.

Vietnamese is a typical Lexical language. In this language all grammatical morphemes are expressed by independent words. The independent grammatical words have similar phonological composition with lexical morphemes. Lexical morphemes in Vietnamese formally coincide with a word and are commonly monosyllabic. Zubkova (1990) says that in Vietnamese consonant combinations in C_1VC_2 words have two main characteristics:

(67) *Restrictions on C_1 and C_2 in C_1VC_2 words in Vietnamese*
- a) $C_1$and C_2 form a rising sonority contour.
- b) C_1 and C_2 belong to different phonemic sets.

Thus, in Vietnamese the Sonority Sequencing Principle works on the word domain and consonants form separate sets depending on the positions in a word: initial and final. For instance, voiced fricatives occur only in word-initial position, while semivowels occur only in word-final position. These patterns are illustrated by the words in (68).

(68)
vay[1]	'to take'	
chay[6]	'to run'	
sâu[2]	'grief'	
gâu[5]	'bear'	(Zubkova 1990:92)[26]

Zubkova points out that there are three parameters according to which C_1 contrasts with C_2: (i) coronal vs. labial/velar, (ii) fricative vs. stop and (iii) obstruent vs. sonorant. The latter type of contrast is the most common.

[26] Superscripts denote tone.

Turkish is a typical Lexico-grammatical language. Lexical morphemes occur independently. Grammatical morphemes can also occur as independent words, but commonly attach to lexical morphemes as suffixes. Lexical and grammatical morphemes differ in their phonological makeup. Similarly to Vietnamese, C_1 and C_2 form a rising sonority contour in C_1VC_2 words in Turkish. Unlike Vietnamese, it is possible in Turkish for both C_1 and C_2 to be obstruents or sonorants. Thus there is no strict differentiation in the occurrence of consonants in initial and final positions. The main parameter according to which C_1 contrasts C_2 is anterior vs. posterior. The words in (69) illustrate all of these patterns.

(69) kap ‘vessel’
gün ‘day, sun’
kol ‘arm’

Russian is a Grammatical language with predominantly dependent lexical and grammatical morphemes. In this language the difference in the phonological makeup of lexical and grammatical morphemes is much greater than in Turkish. Only 14 phonemes (33% of the whole phonemic inventory of Russian) are attested in inflectional morphemes (Abdaljan 1978). Both prefixation and suffixation are common. Lexical morphemes coinciding with a word also occur, e.g. words with a C_1VC_2 structure are attested. Some researchers posit a zero morphemes at the end of such morphemes, e.g. in the case of nouns a zero morpheme stands for the nominative case and gender marker (Kasevič 1986). Consonant co-occurrence in C_1VC_2 words has interesting patterns. Unlike Vietnamese and Turkish, in Russian C_1 and C_2 often form a falling sonority contour. In addition, Russian differs from both Vietnamese and Turkish in that it allows combinations where C_1 is either /r/ or /l/ and C_2 is a stop, combinations which are dispreferred in both Vietnamese and Turkish, e.g.

(70) luk ‘onion’
les ‘forest’
rot ‘mouth’
rak ‘crayfish’

In Russian the main parameter according to which C_1 contrasts with C_2, is coronal vs. labial/dorsal.

To summarise, there are gradual patterns to be observed in these languages. Vietnamese and Turkish are similar in that lexical morphemes are independent in both languages, they coincide with a word and are characterised by rising sonority. In Russian, lexical morphemes can have a falling sonority, perhaps because they very rarely occur independently. As mentioned earlier, Russian has rich prefixing and suffixing morphology. On the other hand, consonant co-occurrence restrictions in these languages show that Vietnamese stands opposed to Turkish and Russian, in that consonants within words in Vietnamese are highly contrastive (three parameters), while in Turkish and Russian consonants contrast only on one parameter. The

greater the contrast is between the C_1 and C_2, the fewer the co-occurrence restrictions between them. Thus Vietnamese, unlike Turkish and Russian, does not have co-occurrence restrictions within a lexical morpheme/word. I think there should be a reason for this asymmetry. The answer to the puzzle could be as follows: lexical morphemes must have formal autonomy in order to be recognisable. The preference for lexical morpheme to retain its form is attested cross-linguistically both by synchronic processes and by language change. Thus, the principle at work in all of these languages is the same: to maintain the formal autonomy of lexical morphemes. The strategy employed to achieve this goal in Vietnamese differs from those employed in Turkish and Russian.

In Vietnamese, in order to achieve the autonomy, the so-called boundary signals are used by way of assigning different sound sets to C_1 and C_2, respectively. Thus, external means are used in Vietnamese; while in Turkish and Russian –because both languages have affixation (suffixation in Turkish and prefixation and suffixation in Russian) – co-occurrence restrictions are employed in order to achieve formal autonomy of lexical morphemes by internal means. These segmental patterns might be comparable with suprasegmental means of demarcation (boundary signals), e.g. a fixed word accent can be considered as an external means which identifies the beginning or the end of a word unequivocally; on the other hand, vowel harmony might be considered as an internal means of word demarcation.

In terms of the phonological make-up of grammatical and lexical morphemes it has been shown that independent though these morphemes are, they are similar in their phonological composition. The pattern gradually changes from Lexical to Grammatical languages. In the latter type grammatical morphemes have only a subset of phonemes occurring in lexical morphemes.

Georgian is a Grammatical language. Lexical as well as grammatical morphemes do not occur independently. From the patterns that have been established above, the following can be expected:

(71) a) Many consonant co-occurrence restrictions within the stem domain.
b) The asymmetry between lexical and grammatical morphemes in terms of their phonological makeup.

Both of these patterns are attested in Georgian (see Chapter 3).

Another point which is related to the interplay between the morphological typology and phonotactics concerns the correlation between lexical and grammatical morphemes in terms of their length (calculated in terms of the number of syllables), which is distinct for each language group. I have already mentioned that lexical morphemes are generally longer than grammatical ones. Here I would like to be more precise and say that, in fact, such correlation largely depends on the type of language. In Lexical languages, the length of lexical and grammatical morphemes is almost the same. This might be related to the fact that in these languages both lexical and grammatical morphemes occur as independent words. The difference between the lexical and grammatical morphemes in terms of length increases from Lexico-

grammatical to the Grammatical languages. Consider the statistical data discussed in Zubkova (1990:169).

(72) *The length of lexical/grammatical morphemes in languages of different types*

Morpheme type / Language type	Lexical	Grammatical
Lexical Yoruba (Bade 1986)	1.75	1.17
Lexico-grammatical Maratxi (Bade 1986)	2.3	1.4
Grammatical Russian (Popova 1989)	2.7	1

It is apparent from table (72) that the difference between the lexical and grammatical morphemes in terms of length (syllable constituency) increases from Lexical to Grammatical languages.

To conclude, the autonomy of lexical and grammatical morphemes varies in Lexical, Lexico-grammatical and Grammatical languages, and affects the phonological composition of a word.

There is another factor playing a role in the phonotactic structure of a word in Lexico-grammatical and Grammatical languages. This concerns the place of the affix in a word; depending on the language, whether it is predominantly prefixing, suffixing or infixing, or has both prefixation and suffixation. Since the thesis does not deal with derived words, the effect is discussed only briefly, in section 2.3.2.1.

2.3. Principles

2.3.0. Introduction

There are a number of phonological principles governing co-occurrence restrictions which bear directly on the domain problem and determine the phonotactic structure of a word.

Greenberg's (1978) article identifies three principles as being the most important for the study of cross-linguistic characteristics of consonant clusters: the Obligatory Contour Principle (OCP), the Sonority Sequencing Principle (SSP) and the Principle of Resolvability (PR). Each principle is defined briefly below.

a) The claim that "… phonemes differentiated by a mark of correlation never combine in the same morpheme" is originally due to Trnka (1936). The claim was later elaborated as the principle of the dissociation of like consonants (Krupa 1967) and as the Obligatory Contour Principle (OCP) (see Leben 1973, 1978 and McCarthy 1986).

(73) *The Obligatory Contour Principle* (OCP) (Leben 1978)
Adjacent identical tones are banned from the lexical representation of a morpheme.

The principle was later extended to identical segments and to distinctive features. It should be noted that all three versions of the principle are defined exclusively in the monomorphemic context.

b) Greenberg's (1978) generalisations on word-initial and word-final consonant clusters provide some support for the Sonority Sequencing Principle (SSP).

(74) *The Sonority Sequencing Principle* (SSP)
The order of segments in a syllable is such that the sonority of segments rises towards the peak vowel and falls from the peak vowel to the end of the syllable.

The principle was originally proposed by Jespersen (1904) and later recognised either as an absolute condition or simply as a preference condition expressing universal markedness values (Blevins 1995). Related to the principle is the highly debated concept of the sonority hierarchy: obstruents > nasals > liquids > glides > vowels (see Clements 1990). The hierarchy depicts the preferred linear order of segments in a syllable, often violated by languages exhibiting long consonant sequences, e.g. Berber, Polish, English, Georgian, etc.

c) The Principle of Resolvability (PR), proposed by Hjelmslev (1936), states:

(75) *The Principle of Resolvability* (PR)
In general, longer consonant sequences contain shorter ones as partial sequences.

I address each of these principles in the following sections. The traditional formulation of the OCP is assumed and substantiated by additional language data. I propose to re-examine the SSP by defining it at the stem or word level, instead of the syllable level, as claimed in previous studies. I also suggest the unification of the SSP with a closely related phenomenon, namely the Syllable Contact Law (SCL). In order to substantiate the proposal, I discuss the typological data of Zubkova (1990). Phonotactic generalisations based on this typological study illustrate that both the SSP and the SCL function within the stem/word domain and reflect the rising and falling sonority pattern of a word. I also propose the extension of the PR to two-member consonant clusters. The reconsideration of the principle leads to a hypothesis about the derived nature of consonant clusters. In addition, it is proposed that the patterns of the clusters are derived from the patterns of the stem. The hypothesis is tested in Chapter 3. Finally, I suggest that all these principles, the OCP, the SSP, the SCL and the PR, are different instantiations of a more general Balancing Principle (BP), defined as follows:

(76) *The Balancing Principle* (BP)
The Balancing Principle is a cover term for the phonological principles referring to well-formedness constraints implying ease of articulation and perception.

The above-mentioned principles are cross-linguistically attested. In Chapter 3 I consider another generalisation, the Compensatory Principle (CP) (originally formulated by Melikishvili 1979, 1997), which seems to be language-specific (see Chapter 3).

Thus, the proposal advocated in this section is the following. The OCP, the SSP, the SCL and the PR are different instantiations of the Balancing Principle (BP). All the principles refer to the universal principles of ease of articulation and perception. They are defined at a single domain, the stem or the word, depending on the language type.

2.3.1. The Obligatory Contour Principle

Two different effects have been attributed to the Obligatory Contour Principle in the Generative framework tradition.

(77) a) *Passive interpretation*: It has been interpreted as a morpheme-structure constraint (Leben 1973, McCarthy 1986, Hyman 1987 and Myers 1987).
b) *Derivational interpretation*: According to this interpretation, the application of a phonological rule is blocked if it creates a violation of the OCP (McCarthy 1986, Archangeli 1986, Itô & Mester 1986, Borowsky 1987, Myers 1987 and Yip 1988). The OCP has also been interpreted as a rule trigger (Yip 1988).

The discussion that follows concerns only the first application of the OCP. Some derivational facts of Georgian referring to the OCP are considered in Chapter 3.

Since the thesis focuses primarily on consonant phonotactics in general, and consonant clusters in particular, the data that follow mainly concern the effects of the OCP on consonants. Two interesting observations made in almost all studies discussed in this section (Chrétien 1965, Krupa 1971) and the next (Zubkova 1990) are:

(78) a) Vowel co-occurrence displays an associative tendency.
b) Consonant co-occurrence displays a dissociative tendency.

The co-occurrence of sounds with the same place of articulation, manner or laryngeal specification is called associative, while the combination of dissimilar sounds is dissociative. To re-state the generalisation, vowels tend to have associative combinations e.g. combinations of the vowels *a-a* are attested in many languages, e.g. Czech *kabát* ‘coat’, Tadjik *navad* ‘ninety’, Russian *saxar* ‘sugar’, Japanese *takara* ‘treasure’, Javanese *tawang* ‘air’, Tagak *parang* ‘forest’. Consonant combinations

tend to be disharmonious. The principles of the OCP and the SSP substantiate this generalisation. The fact that vowel harmony is cross-linguistically a more widespread process than consonant harmony could be attributed to this observation.

The generalisation in (78) is remarkable in that it bears on the correlation mentioned earlier between the association of consonants with the lexicon and vowels with the grammar (see Chapter 1). Furthermore, the OCP, the SSP, the SCL are dissociative in nature.

The OCP constrains the occurrence of consonants with similar (i) place of articulation, (ii) manner and (iii) laryngeal specification. A language may have all of the listed types of OCP constraints or only one of them.

Consider, for example, consonant co-occurrence restrictions in eight Polynesian languages: Hawaiian, Tuamotuan, Maori, Rarotogan, Tahitian, Easter Island, Ceremonial Samoan and Tongan. Krupa's (1971) study reports the existence of the OCP constraint on place of articulation within the lexical morpheme in all of these Polynesian languages.

(79)

Language C1-C2	East. Isl.	Haw.	Tah.	Tuam.	Rar.	Maori	Sam.	Tong.
[labial] - [labial]		–	–	–	–	–	–	–
[labial] -[coronal]						+	+	+
[labial] - [dorsal]								
[coronal] - [labial]				+		+	+	+
[coronal] – [coronal]			–	–	–	–	–	–
[coronal] - [dorsal]								
[dorsal] - [dorsal]		+						
[dorsal] - [coronal]								
[dorsal] - [dorsal]								

Three sets of consonants are distinguished: (i) [labial], (ii) [coronal] and (iii) [dorsal].

The symbol '+' in (79) denotes that the combination is attested, e.g. the combinations of bilabials and coronal consonants ([labial] - [coronal]) are attested in Maori, Ceremonial Samoan and Tongan. The symbol '–'denotes that combinations of consonants are unattested, e.g. combinations of two bilabials ([labial] - [labial]) consonants are unattested in Hawaiian, Tuamotuan and Maori. The blank space in the table denotes that combinations are not expected, on a language-specific basis.

According to Krupa (1971), the Polynesian languages cited in (79) do not have consonant clusters. Thus, the constraints on a place of articulation hold across a vowel within the lexical morpheme. As seen from (79), the languages disallow combinations of consonants with the same place of articulation, e.g. of the types [labial] - [labial] or [coronal] - [coronal]. According to Krupa (1971), the most common combinations of consonants in these languages are of the types [labial] - [coronal] and [coronal] - [labial].

The OCP is also evident in Semitic languages. Various types of OCP effects, including constraints on place of articulation and manner of articulation, are attested in Proto-Semitic roots. Consider some of the generalisations proposed by Greenberg (1950).

(80) a) Outside of /w/ and /y/, there are four sections of consonants: back consonants /ʔ h ḥ q x ɣ k g q/; liquids /r l n/; front consonants /ḍ š ś s z ṣ t d ṭ ṯ ḏ ṱ/ and labials /p b m/. Consonants of any one section occur freely with those of any other section in the formation of tri-consonantal verbal morphemes.

b) Different consonants of the same order tend not to appear in the same tri-consonantal verb morphemes. The rule has some exceptions for the velars, the pharyngeals and the laryngeals.

c) There are no Proto-Semitic roots with identical consonants in the first and second positions. On the other hand, identical second and third consonants are very common. This explains the non-occurrence of geminate consonants in word-initial position and their appearance in the word-medial one.

d) The above statements only apply to the verb root morphemes. Substantival morphemes frequently violate them.

(Greenberg 1950:178)

Many researchers considered remarkable the fact that the OCP constraints are primarily operative on the first and the second consonant of tri-consonantal verbal roots. This observation led many to claim that Semitic roots are primarily biconsonantal, and that the third consonant has a grammatical origin (Paradis & Prunet 1993 and Bohas 1997). The issue is beyond the scope of the discussion here, but is mentioned to emphasise once again how phonotactics and morphology are interrelated.

The data on OCP constraints referring to place of articulation and manner of articulation have been considered. The OCP can also refer to a laryngeal feature. For instance, Japanese has a constraint against the occurrence of two separate voiced obstruents within a morpheme (Itô & Mester 1986).

OCP constraints referring to place of articulation, manner and laryngeal features are operative in Georgian lexical morphemes and are discussed in Chapter 3.

Various applications of the OCP (although the term was not used as such in early works) have been reported for many other languages, e.g. Czech (Mathesius 1929), English (Trnka 1936), German (Twaddell 1939, 1941), Afrikaans (Odendal 1962), Italian (Krámský 1964), Swedish (Sigurd 1965) and Arabic (Krupa 1967). The data are numerous and I will not go into any further details. The main point is to establish that the Obligatory Contour Principle constrains the co-occurrence of consonants in various ways and operates exclusively within the lexical morpheme domain.

2.3.2. The Sonority Sequencing Principle

2.3.2.0. Introduction

Traditionally, it has been claimed that the domain of the SSP is the syllable (Clements 1990). Selkirk (1984a) claims that the SSP "… can be viewed as imposing universal constraints on the possible form of language-particular sets of conditions on syllable structure" (Selkirk 1984a:117). Related to the principle is one which was originally proposed by Hooper for Spanish, and later generalised as the Syllable Contact Law (Murray & Vennemann 1983). The law states:

(81) *The Syllable Contact Law* (SCL)
Sonority of a syllable-final consonant must exceed that of a following syllable-initial consonant (equivalently, the second must exceed the first in "strength").

In the following sub-sections I will demonstrate that the SSP and the Syllable Contact Law are in fact different instantiations of a single phenomenon. Instead of the syllable domain, both the SSP and the SCL should be defined at the lexical morpheme domain, or the word domain, depending on the language type. Both generalisations are artefacts of the rising and falling sonority contour of a word. In the following sub-section I discuss the typological study by Zubkova (1990). Zubkova's data support the claim that both principles refer to a single phenomenon, which is a characteristic feature of the lexical morpheme/word, not of the syllable.

2.3.2.1. Statistical data from Zubkova (1990)

Zubkova (1990) reports and discusses the results of a study on the phonological typology of the word, based on eleven languages which differ in their genetic affiliation and morphological type. The study focuses on general tendencies of co-occurrence restrictions on vowels and consonants within words of the CVCVCV type and stems of the CVC type. It mainly concerns the co-occurrence restrictions on the primary vowels /i a u/ and the primary consonants /p-b t-d k-g m n s (r l)/.[27] I consider here the data on the co-occurrence patterns of consonants.

First, the languages examined in the study are introduced.

[27] The distinction between primary and secondary sounds was first proposed by Jakobson (1941) and later extended by Milewski (1967, 1969) and Skalička (1970).

(82) *Language data*[28]

Languages	Language Family	Morphological Type	Type of Affix	Word	Accent/prosody
Russian	Indo-European	Grammatical	Prefix, suffix	Stem + affix	Free
Czech	Indo-European	Grammatical	Prefix, suffix	Stem + affix	Fixed on the first syllable
Tadjik	Indo-European	Lexico-grammatical	Suffix	Stem	Final fixed
Indonesian	Malay-Polynesian	Lexico-grammatical	Prefix	Stem	Stem-level vowel harmony
Javanese	Malay-Polynesian	Lexico-grammatical	Prefix	Stem	Stem-level vowel harmony
Tagalog	Malay-Polynesian	Lexico-grammatical	Prefix	Stem	Free
Turkish	Altaic	Lexico-grammatical	Suffix	Stem	Final fixed
Mongolian	Altaic	Lexico-grammatical	Suffix	Stem	Fixed on the first syllable
Nanai	Altaic	Lexico-grammatical	Suffix	Stem	Weak, final
Japanese	Altaic/ Austroasiatic	Lexico-grammatical	Suffix	Stem	Free
Vietnamese	Austroasiatic	Lexical	No affix	Stem	Tone

According to Zubkova (1990), in all these languages a general tendency can be observed within a lexical morpheme of the C_1VC_2V type, i.e. that the sonority of the C_1 is less than the sonority of the C_2. The generalisation is depicted in (83).

(83)
$$\begin{array}{c} C_1VC_2V \\ | \quad | \\ \text{obstr.} \quad \text{sonorant} \end{array}$$

The data illustrate that the SSP is operative across a vowel within a lexical morpheme/word. Interestingly, exceptional cases are attested in prefixal languages (e.g. Tagalog) where the typical C_1 is a sonorant consonant. The observation gives a hint as to a possible effect which the type of affixation (prefixation or suffixation) might have on the phonotactic structure of a word. For example, prefixing languages often violate the SSP, e.g. Russian (Zubkova 1990) and Zoque (Booij 1983). The same effect is attested in Georgian, which will be considered in Chapter 3.

The following discussion is devoted to this interplay between the type of affixation and consonantal patterns of a word.

[28] In order to be consistent with the previous discussion on morphological typology, I have replaced Zubkova's original morphological types with the new terminology: Lexical, Grammatical and Lexico-grammatical, respectively. The following replacements have been made: Inflectional is changed into Grammatical; Analytical, Agglutinative and Agglutinative-Analytical are replaced by Lexico-grammatical; and Isolating is replaced by Lexical.

Tables (85) and (86) summarise the results of calculations on the sonority profile of the consonants in the words of $C_1VC_2VC_3V$ type in terms of two parameters: the laryngeal and manner features. The pattern is studied in two types of languages: prefixal and suffixal. Consider the following examples, from a suffixal language, Turkish, and from a prefixal language, Javanese. Both languages belong to the Lexico-grammatical language group, i.e. have predominantly agglutinative type of morphology.

(84) Turkish /kavun/ 'melon' /torun/ 'grandchild'
Javanese /buluk/ 'dirty from dust' /sorot/ 'beam, ray'

The first consonant (C_1) in Turkish is [–voiced] and the second and the third consonants (i.e. C_2 and C_3) are sonorants. In Javanese, the first consonant (C_1) can be either [+voiced] or [–voiced], while the second is occupied by a sonorant and the third by a [–voiced] consonant.

The general tendency of the occurrence of consonants in $C_1VC_2VC_3V$ words, attested by Zubkova (1990), is depicted in (85).

(85)

C_1; C_2; C_3 Languages	C_1	C_2	C_3
Suffixal	[–voiced]	[sonorant]	[sonorant]
Prefixal	[–voiced] [+voiced]	[sonorant] [+voiced]	[–voiced]

In terms of the laryngeal features, suffixal languages are characterised by a rising sonority contour throughout the whole word, while prefixal languages are characterised by a rising and falling sonority contour. In both types of languages, the biggest sonority contrasts are attested within the stem, i.e. the lexical morpheme domain; in the case of suffixal languages between C_1 and C_2, and in the case of prefixal languages between C_2 and C_3.

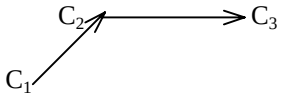
The results of the calculations for manner feature are given in (86).

(86)

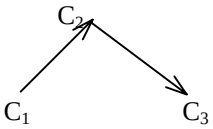
C_1; C_2; C_3 Languages	C_1	C_2	C_3
Suffixal	stop	sonorant	sonorant
Prefixal	sonorant fricative	sonorant stop fricative	stop fricative

The results in tables (85) and (86) are similar. Suffixal languages are characterised by a rising sonority contour along the whole word, while prefixal languages by a rising and falling sonority contour.

(87) a) Suffixal (Turkish) rising sonority



b) Prefixal (Javanese) rising and falling sonority



Additionally, tables (85) and (86) demonstrate that in both types of languages the consonantal inventory of the stem domain is richer than of the affix domain.

(88) a) Suffixal (Turkish) $[[C_1VC_2]_{stem}[VC_3]_{suffix}]_{word}$
$C_1 \leftrightarrow C_2 \approx C_3$
stem = $C_1 \leftrightarrow C_2$

b) Prefixal (Javanese) $[[C_1V]_{prefix}[C_2VC_3]_{stem}]_{word}$
$C_1 \approx C_2 \leftrightarrow C_3$
stem = $C_2 \leftrightarrow C_3$

The symbol '↔' denotes a dissociative relation and '≈' denotes an associative relation.

The asymmetric patterns of consonant distribution in prefixal and suffixal languages explain the cross-linguistic tendency in favour of suffixation as opposed to prefixation. Languages prefer to make use of contrasts at the very beginning of a word rather than the end. This observation has been justified in many instances by experiments on language perception and production (Marslen-Wilson & Zwitserlood 1989).

There are also languages either with no affixation or with affixation of both types. Consider a case from Vietnamese, a Lexical language, and Russian, a Grammatical language. The patterns of these languages are as follows.

(89) a) Languages with no affixation Vietnamese $[[C_1VC_2]_{stem}]_{word}$
$C_1 \leftrightarrow C_2$

b) Languages with both types of affixation Russian $[[C_1VC_2]_{stem}[VC_3]_{suffix}]_{word}$,or
$[[C_1V]_{prefix}[C_2VC_3]_{stem}]_{word}$
$C_1 \approx C_2 \approx C_3$

Vietnamese words are characterised by the occurrence of certain phoneme classes only in certain positions of a word: initial, medial or final. For instance, the obstru-

ents *th* and *kh* are typical initial consonants (i.e. C_1), e.g. *thi* 'examination', *tha* 'set free', *khi* 'when' and *khe* 'slit, chink'; the typical word-medial consonants are glides, while the consonants *p* and *c* are the typical final ones (i.e. C_3), e.g. *dáp* 'to reply', *sóc* 'squirrel' and *học* 'study' (Đình-hoà 1997, Chù & Nguyệt 1997). The typical word-initial nasal is /m/, while the typical word-final is the nasal /n/ (Zubkova 1990:115–116). Thus, in general terms, one can conclude that Vietnamese $C_1VC_2VC_3$ words, in which all consonants belong to the lexical morpheme/word, are characterised by a rising and falling sonority contour. Actually, Lexical languages demonstrate most obviously that the domain of both the SSP and the SCL is the word. This is mainly due to the fact that in Lexical languages grammatical affixes do not attach to the lexical ones, and consequently do not change the 'original' phonotactic shape of a word, which has a rising and falling sonority contour.

Russian, with both types of affixation, presents the opposite pattern from Vietnamese, i.e. with no phoneme sets associated to the position in a word. Nevertheless, Zubkova (1990) points out that there is a preference for a rising sonority contour within the words of the C_1VC_2V type in Russian.

2.3.2.2. Conclusions

To conclude, the consonants within a word are not distributed randomly, but their contrastive function is used in various ways, depending on the type of affixation. Most consonantal patterns are attested within the lexical morpheme domain, which sometimes coincides with the word (e.g. in Vietnamese) and sometimes does not (e.g. in Russian). This means that the lexical morpheme is the optimal domain for studying consonantal patterns of a language.

Common to all types of languages is a general tendency for a rising and falling sonority contour within a word of the $C_1VC_2VC_3$ type. This means that the domain of the SSP is C_1VC_2V and the domain of the SCL is C_2VC_3. One can unify both principles and say that both the SSP and the SCL function on one domain, the stem/word[29] and that they are instantiations of a single phenomenon: the rising and falling sonority contour of a word.

2.3.3. The Principle of Resolvability

The phonotactic study of any language comprises the investigation of constraints not only of adjacent sounds, but also those of distant ones. Co-occurrence restrictions are found in both cases, and interestingly, are very similar. The correlation is formulated as a one-way implication in (90).

(90) Stem: CV << CVS << CVC
Cluster: C << CS << CC
C = obstruent; S = sonorant; V = vowel
<< denotes entailment

[29] Booij (1992), based on Dutch data, also claims that the domain of the SCL is the prosodic word.

The implication states that the existence of a CVS-type stem structure presupposes the existence of a CV-type stem structure, and that the existence of a CVC structure presupposes the existence of a CVS structure. The implication described for stems holds for clusters as well. The existence of an obstruent cluster presupposes the existence of cluster including a sonorant. This observation was borne out by the examination of the distributional regularities of consonant clusters in fixed accent languages (Butskhrikidze 1998a). For example, there are word-initial CS but not CC obstruent clusters in Armenian and in French; Armenian does not have word-initial obstruent clusters except those of the obstruent + sonorant type, e.g. the obstruent + glide /j/. The Armenian form is taken from Vaux (1998).

(91) *Armenian word-initial consonant sequences*

bjur	'10,000'

Examples of consonant + sonorant combinations in French come from Schane (1968) and Charette (1990).

(92) *French word-initial consonant sequences*

travail	'work'
bretelle	'braces'
crevaison	'puncture'

Exceptions to the generalisation are clusters of the /s/ + C type in both languages (e.g. *skizb* 'beginning' in Armenian or *stylo* 'pen' in French). /s/ + stop clusters are generally considered to behave in a very particular way (Kuryłowicz 1975, Fudge 1969, Selkirk 1982 and van der Hulst 1984), and in some studies are treated as complex segments (see van de Weijer 1996). Thus, Armenian and French are not exceptions for the implication CS >> CC.

The parallel patterns found within a stem and consonant clusters lead to a reconsideration of the Principle of Resolvability; namely its extension to the two-member clusters, and its correlation to the stem domain. The claim is that all two-member consonant clusters are derived,[30] the result of vowel deletion. Their characteristic patterns and co-occurrence restrictions are derivable from stems of the CVC type. The proposal is formulated as a hypothesis in (93).

(93) *Hypothesis*
If a language has clusters of the C_iC_j type, then the language will have stems of the C_iVC_j type.

The merit of this hypothesis is that it is easily refutable, is easily testable and has predictive power. The plausibility of the implication will be tested on Georgian data in Chapter 3.

[30] The term *consonant cluster* refers to a genuine cluster, not to one which is the result of complex segment formation, or of morphological merger, i.e. the result of a lexicalisation process.

2.3.4. Phonotactic generalisations

This section concerns some phonotactic patterns: the tendencies for certain sounds to occur in certain positions of a word. The generalisations are cross-linguistically attested. I do not have any explanations to offer for these tendencies at this point. I would merely like to list briefly some of the generalisations, and leave the issue open for further study. Interestingly, the tendencies listed below are also attested in Georgian.

The first observation concerns the distribution of bilabials. As Zubkova (1990) and Chrétien (1965) report, there is a tendency for bilabials to occur word-initially. The occurrence of bilabials in word-final position is very restricted. This generalisation brings to mind my observation that bilabials commonly appear as the first member of long consonant sequences, e.g. in Polish and in Georgian. The asymmetric distributional characteristics of the nasals /m/ and /n/ could be attributed to this generalisation. The nasal /m/ is commonly found in word-initial position, while the nasal /n/ commonly occurs in word-final position (Zubkova 1990). The asymmetry is confirmed by the neutralisation of the opposition /m/ vs. /n/ in word-final position in a number of languages, where only the nasal /n/ is attested, e.g. in Old Greek, Italian, Finnish, Avar, Lak and Japanese (Zubkova 1990:108). These restrictions certainly make bilabials more *marked* in many instances, as has been argued by many linguists (see Melikishvili 1976).

The second observation concerns the asymmetric patterns of the sonorants /r/ and /l/. There is a strong tendency for the sonorant /r/ not to appear in word-initial position, while such a restriction does not apply to the sonorant /l/ (Zubkova 1990). The asymmetry finds its expression in the dissimilar combinatory patterns of these liquids in clusters (see also van der Torre 2000).

Although I avoid speculating about possible explanations for these generalisations, I would like to say that it would not surprise me if future studies show that these generalisations are in some sense instantiations of the Balancing Principle.

2.3.5. Conclusions

To summarise, the section has addressed phonological principles, their characteristics and the domains within which they function. The principles of the OCP, the SSP and the PR have been discussed. The traditional formulation of the OCP has been assumed and substantiated by additional language data. I have offered a re-examination of the SSP by defining this principle on the stem/word domain, instead of the syllable, as claimed in previous studies. In addition, I have suggested the unification of the SSP and the closely related Syllable Contact Law (SCL). In order to substantiate the proposal, I have discussed the typological data given by Zubkova (1990). The generalisations illustrated that both the SSP and the SCL function within a stem/word domain and reflect the rising and falling sonority pattern of a word. Furthermore, during a re-examination of the PR I suggested the extension of the PR to two-member consonant clusters. The reconsideration of the principle led to a hypothesis about the derived nature of consonant clusters. Additionally, it has been

proposed that the patterns of the clusters are derived from the patterns of the stem. This hypothesis is tested in Chapter 3. Finally, I suggested that the OCP, the SSP, the SCL and the PR are different instantiations of the Balancing Principle (BP). All of them attempt to achieve universal requirements of language: ease of articulation and ease of perception. They are defined on the stem, or the word, depending on the language type.

2.4. General conclusions

Form is characterised by discreteness, meaning by hierarchy. The phonotactics of a word imply the study of the formal (discrete) parts, but also of the meaningful units (morphemes). Thus, notion of *hierarchy* comes into play. A definition of hierarchy is formulated in (94).

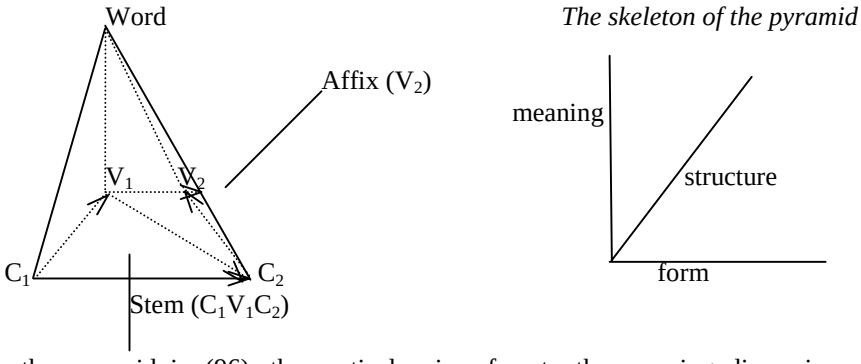
(94) *Definition of hierarchy*
A hierarchy is a ranked structure, the arrangement of a classification scheme from general to specific.

The definition of hierarchy brings up another notion, namely that of structure.

(95) *Definition of structure*
Structure is the way something is put together. It is the pattern of organisation of the whole.

Thus, word phonotactics encompasses three simultaneously interactive components. These components are meaning, form and structure. The word structure can be formalised as a three-dimensional pyramid, as depicted in (96).

(96) *Word phonotactics*

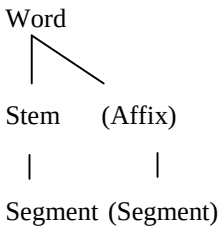


In the pyramid in (96), the vertical axis refers to the meaning dimension, the horizontal axis refers to the formal, segmental substance (i.e. phonetic) dimension

and the diagonal axis refers to the structural dimension. The latter includes the general organisational principles, e.g. the phonological principles.

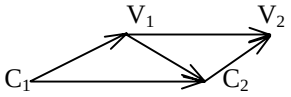
The pyramid includes three hierarchically organised units: segment – stem – word:

(97)



A particularly striking feature of the hierarchy is that an increase in complexity is coupled with the emergence of new characteristics. Since segments' (phonemes') main function is distinctive, they are already incorporated in the morphemes in the proposed representation (see (96)). Unlike previous approaches, where every segment was required to be a part of the syllable, my claim is that every segment is necessarily a part of the morpheme (whether lexical or grammatical). The morphemes are integrated within a higher domain, the word, which stands between morphology and syntax and has distinct phonological and morphological characteristics. As a general word-frame I assume the disyllabic $C_1V_1C_2V_2$ word (McCarthy & Prince 1986b). The word is characterised by a rising and falling sonority pattern, as depicted in (98).

(98)



In addition to the sonority contour of a word, the diagram in (98) depicts the adjacency and precedence relations between the segments: C_1, V_1, C_2 and V_2. C_1 precedes C_2 and is adjacent to it. Adjacency entails the existence of a certain relation (in this case co-occurrence restrictions), which holds between the adjacent segments. Indeed, the phonological principles, the OCP and the SSP, are operative on the consonants C_1 and C_2, which are part of the stem, and ultimately of the word domain. As seen from the diagram, the segments C_1 and V_2 are not adjacent, and consequently the existence of co-occurrence restrictions is not expected between them. In addition, the diagram illustrates that consonant clusters are potentially derivable from the C_1VC_2 stem after deletion of the stem vowel, and that two-member consonant clusters are the most widely attested type of cluster cross-linguistically. These clusters obey phonological principles such as the OCP and the SSP. Surface clusters which violate these principles are expected to be the result of a morphological merger, e.g. of a prefix with a stem, as it is the case in Georgian (see Chapter 3). It is sometimes difficult to demonstrate that such synchronic surface clusters are morphologically

derived, but in most of the cases it is nevertheless possible to trace their origin by taking into consideration synchronic, comparative and historical evidence, as suggested in Chapter 3 and Chapter 6.

In the following chapter the Georgian phonotactic data are presented. The hypothesis formulated in (93) is verified. It is followed by an analysis of consonant clusters, the Gradual Consonant Analysis (see Chapter 6), which substantiates the plausibility of the representation of the word proposed in (96).

3 Georgian Phonotactics

3.0. Introduction

In this chapter, I present extensive descriptive data from the phonotactics of Georgian. This is meant as background to the assumptions that I make in the analysis of Georgian phonotactics, specifically the analysis of Georgian consonant sequences in Chapter 6.

The study is restricted to Modern Literary Georgian. A more comprehensive study would be needed to account for richer dialect data, e.g. incorporating vowel systems with length distinctions and umlaut. Nevertheless, dialect data are sometimes discussed when relevant for the justification of the plausibility of particular arguments.

This chapter is organised as follows: in 3.1 a general introduction to the Georgian language is given. Morphological patterns are discussed with respect to the phonotactic structure of a word; in 3.2 the phonemic inventory and general syntagmatic regularities of vowels are considered; in 3.3 special attention is paid to the consonantal system, represented by three sets: stops, fricatives and sonorants. Two important claims are made: the sonorant /r/ is syllabic in a consonantal environment, and combinations of C + /v/ are complex segments; in 3.4 specific characteristics of the Georgian minimal word are discussed. The disyllabic minimal word constraint is proposed; in 3.5 a general introduction to consonant syntagmatics is offered. All possible combinations of consonants of different lengths and constituency are discussed. Harmonic clusters are argued to be complex segments, on the basis of distributional, comparative and historical evidence. It is proposed that the longest true consonant sequences are biconsonantal; in 3.6 the hypothesis concerning the derived nature of consonant sequences is tested. The restrictions on consonants at a distance (across a vowel) and in adjacency (in sequences) are compared; in 3.7 the generalisations are summarised and general conclusions on Georgian phonotactics are drawn.

3.1. The Georgian language

Georgian belongs to the Kartvelian (South Caucasian) language family. The name of the language group is related to the ethnonym *kartveli* ‘Georgian’. Georgian is the official state language of Georgia, with more than 5 million speakers. Outside Geor-

gia, there are Georgian-speaking populations in Azerbaidjan (Saingilo),[1] Turkey (Shavsheti, Imerkhevi), Iran (Fereidan) and the North Caucasus (Sochi, Kizlar-Mozdok, Orjonikidze). Besides Georgian, the Kartvelian language family consists of three other languages, Megrelian (Mingrelian), Laz (Chan) and Svan. The Kartvelian languages are spoken within the territory of Georgia, the Central and West Caucasus and the vast territory of the South Caucasus. Megrelian is spoken in the western part of Georgia, namely, in the northern part of the Kolkheti Lowlands and in the neighbouring mountainous region. Two dialects of this language can be distinguished: Senakuri and Zugdidur-Samurzakanuli. Laz is spoken near the Black Sea, from the village of Sarpi (on the southern border of Georgia) to Kerem (Turkey). It consists of three dialects: Atinuri, Vicur-Arkabuli and Khopuri. Svan is spoken in the northwest of Georgia. Four dialects of this language can be distinguished: Balszemouri, Balskvemouri, Lashkhuri and Lentekhuri.

The kinship and common origin of the Modern Kartvelian languages has been confirmed by numerous studies on phonetic and morphological correspondences (Güldenstädt 1787, Deeters 1926, 1930, Bopp 1846, 1847, Brosset 1849, Gamkrelidze & Machavariani 1965, Sardschweladse & Fähnrich 1990). Much research has been done to establish the genetic affiliation of Kartvelian languages with other language families or individual languages, e.g. Basque (Bouda 1949, 1950), Indo-European (Bopp 1846, 1847), Semitic (Marr 1908) and North Caucasian (Bouda 1954–56, Chikobava 1965). Although a number of typological similarities have been found, to my knowledge no one has demonstrated beyond doubt the genetic affiliation of the Kartvelian language group to any other language group in the world.

According to Shanidze's (1973) classification, which is mainly based on geographical factors, there are six groups of Georgian dialects: 1. Pkhouri (Khevsuruli, Mokheuri, Tushuri); 2. Mtiulur-Pshauri (Mtiulur-Gudamakruli, Pshauri); 3. Kartlur-Kakhuri (Kartluri, Kakhuri, Javakhuri, Meskhuri, Kizikuri); 4. Dasavluri (Imeruli, Guruli, Rachuli, Lechkhumuri); 5. Samkhret-Dasavluri (Acharuli, Imerkheuli); and 6. Ingilouri.

Georgian is the only written language of the Kartvelian language family. The oldest written documents that are still preserved can be dated back to the fourth century. According to some specialists, the writing system is much older than the texts, and its origin is related to the Georgian king Parnavaz, who reigned in the third century B.C. Gamkrelidze (1989) has argued for a connection between the origin of the Georgian alphabet and the conversion of Georgia to Christianity. Taking into account the paradigmatic and syntagmatic characteristics of the Georgian alphabet, he demonstrates the influence of the Greek alphabet on the Georgian one. The same idea had been expressed by Gardthousen (1879), Bakradze (1889) and K'ek'elidze (1929), but had not been systematically demonstrated. The alloglottographic method[2] of writing was widely used in Georgia and in the Caucasus region

[1] Names of the geographical areas are given in brackets.

[2] The alloglottographic method involves the transcription of a language with symbols of the writing system of another language. The method is commonly used when a language does not have its own writing system.

in general, before the spread of Christianity. Prior to the introduction of the alphabetic system, the Georgian language was encoded through the Aramaic writing system.

The modern Georgian writing system is based on the round-form cursive, which was developed from the angular book script of the ninth century; the latter was a direct descendant of the Old Georgian script. The Georgian writing system includes a symbol, represented by a single grapheme, for each of the distinctive sounds (phonemes) of the Georgian language. The phonemic inventory of Georgian consists of 33 phonemes: 28 consonants and 5 vowels. In the table in (1) all phonemes are presented in alphabetical order. Phonemes are presented in their original Georgian graphic form (i), followed by the name of the symbol (ii), the notation used throughout this thesis (iii) and IPA equivalents (iv). Note that the symbol **'** denotes glottalisation in obstruents.

(1) *The Georgian Alphabet*

(i)	(ii)	(iii)	(iv)	რ	rae	r	[r]
ა	an	a	[a]	ს	san	s	[s]
ბ	ban	b	[b]	ტ	t'as	t'	[t']
გ	gan	g	[g]	უ	un	u	[u]
დ	don	d	[d]	ფ	par	p	[p]
ე	en	e	[e]	ქ	kar	k	[k]
ვ	vin	v	[v]	ღ	ɣan	ɣ	[ɣ]
ზ	zen	z	[z]	ყ	χ'ar	χ'	[χ']
თ	tan	t	[t]	შ	šin	š	[ʃ]
ი	in	i	[i]	ჩ	čin	č	[tʃ]
კ	k'an	k'	[k']	ც	can	c	[ts]
ლ	las	l	[l]	ძ	jil	j	[dz]
მ	man	m	[m]	წ	c'il	c'	[t's]
ნ	nar	n	[n]	ჭ	č'ar	č'	[t'ʃ]
ო	on	o	[o]	ხ	xan	x	[χ]
პ	p'ar	p'	[p']	ჯ	ǰan	ǰ	[dʒ]
ჟ	žan	ž	[ʒ]	ჰ	hae	h	[h]

Before considering the phonemic inventory, a short introduction to Georgian morphology is in order. Georgian morphology is rich, with a very productive inflectional and derivational morphology. Since the technique of morpheme combination is both agglutinative and inflectional, Georgian can be characterised as a Grammatical language (see Chapter 2 for the terminology). Grammatical affixes attach before or after a root, and may sometimes even violate the integrity of the root if a metathesis process takes place. Both the noun and the verb are characterised by prefixing and

suffixing morphology, but verbal morphology is much more complex. Consider the morphologically complex nominal (2a) and verbal (2b) forms in (2).

(2) a) kud-ian-eb-is-tvis ‘for people wearing hats’

kud-	N ROOT
-ian-	POSS
-eb-	PL
-is-	GEN
-tvis	POSS

b) ga-u-k’et-eb-in-eb-i-a ‘somebody made somebody else do something’

ga-	PREV
-u-	PRE-RADICAL VOWEL (OBJECTIVE VERSION MARKER)
-k’et-	V ROOT
-eb-	THEM SUFF
-in-	CAUS
-eb-	THEM SUFF
-i-	PERF
-a	3rd SUBJECT MARKER

When affixes attach to a root, several phonological processes can take place, e.g. vowel deletion, metathesis and epenthesis. The deletion of the root vowel is the most widespread process in both nominal and verbal morphology. As a result of vowel deletion, consonant sequences emerge, as illustrated in (3).

(3) a)

NOM	GEN	
xmal-i	xml-is	‘sword’
cxvar-i	cxvr-is	‘sheep’

b)

da-č’er-i	‘cut IMP’	da-č’r-a	‘he had cut’
PREV - V ROOT		PREV - V ROOT - 3rd MARKER	

Surface consonant sequences also emerge as a result of simply adding consonantal prefixes or suffixes to a root, e. g.

(4) a) gv-c’er ‘you write us’
2nd PL MARKER - V ROOT

b) gv-xat’-av ‘you paint us’
2nd PL MARKER - V ROOT - THEM SUFF

c) m-c’er-al-i ‘writer’
NOMINALISER - V STEM - PRT MARKER - NOM MARKER

Thus, one of the ways to account for the surface consonant sequences of Georgian is to study the morphological structure of such words. Sequences may be the result either of the deletion of a root vowel, which generally happens when a vowel-initial affix is added to a root, as illustrated in (3), or of the addition of a consonantal affix to a root, as illustrated in (4).

I have already mentioned that, in general, lexical and grammatical morphemes differ in their use of the phonological inventory. Morphological languages demonstrate the discrepancy between the two: grammatical morphemes use a subset of the phonological inventory of a language, whereas the full set is employed for lexical morphemes. Georgian, as a Grammatical language, also demonstrates this pattern. I will briefly summarise my findings concerning this issue.

Of the total of 28 consonants of Georgian, 13 can occur in both lexical and grammatical (derivational and inflectional) morphemes; while the other 15 occur exclusively in lexical morphemes.

(5) a) Consonants appearing only in lexical morphemes:
/p p’ t’ k č j c’ č’ ǰ z š ž χ’x h/
b) Consonants appearing in lexical and grammatical morphemes:
/b t d g k’ c s ɣ v m l r n/

It is interesting to note that the set in (5a) includes almost all affricates (i.e. consonants with a complex stop + fricative constituency), e.g. /č c’ č’ j ǰ/, and other marked consonants, e.g. /h/ and /ž/, the distribution of which is quite restricted, e.g. /h/ occurs only in word-initial position, while the fricative /ž/ occurs only in a few words of Georgian origin.

Consonants in grammatical affixes given in the set in (5b) have different distribution patterns depending on their position in a derived word.

(6) a) Word-initial consonants: /v m n s[3] d g/
b) Word-medial consonants: /v m l r n t d g b k’ ɣ/
c) Word-final consonants: /v m n s t d t b[4] c/

It is interesting to observe that the velars /k’/ and /ɣ/ appear only in word-medial position, the affricate /c/, which is the only affricate consonant that appears in grammatical morphemes, occurs only in word-final position and the liquids /r/ and /l/ occur only in word-medial position. Coronals, nasals and the sonorant /v/ appear to be quite free in their distribution. All of this has direct consequences for the surface complexity of consonant sequences in Georgian, and I return to this issue later. One more thing to notice is that of the 13 consonants given in (5b), only eight, /v m n s t d g b/, occur in inflectional affixes, e.g.

[3] It should be noted that depending on the root-initial sound, the person marker /s-/ has several allomorphs, for instance, /h-/ and /š-/. In Old Georgian, the morpheme also had an allomorph /x-/.
[4] Word-final /d/ and /b/ devoice and are pronounced as [t] and [p], respectively, e.g. /k’arg-ad/ ‘well’ is pronounced as [k’argat] and /v-a-k’eteb/ ‘I do’ is pronounced as [vak’etep].

(7) /v-/ 1st SUBJECT MARKER
/m-/ 1st OBJECT MARKER
/s-/ 3rd OBJECT MARKER
/-t/ PL MARKER
/-d-/ POSS MARKER
/g-/ 2nd OBJECT MARKER
/-b/ is attested in the thematic suffix /-eb/

The constituency of the set is reminiscent of the 'primary' consonants discussed in Chapter 2, which appear to be the most unmarked consonants not only in Georgian, but also cross-linguistically. In previous chapters I discussed the correlation between two asymmetries, consonant vs. vowel and lexicon vs. grammar, and proposed that consonants and the lexicon are on the same plane, as opposed to vowels, which are on a par with the grammar. Georgian data also substantiate this cross-linguistic observation. According to Ertelishvili (1970), there are no Georgian lexical roots containing only one vowel; however lexical, monoconsonantal and biconsonantal roots are quite common. For example, Ertelishvili (1970) gives lists of 14 such monoconsonantal and 45 biconsonantal roots. There are also longer roots containing three consonants (42 such roots are attested) or maximally four consonants (14 such roots are attested). I return to the patterns of consonantal roots by discussing their respective histories in Chapter 6. I will propose that such roots are derived; they emerge as the result of vowel deletion, and are sometimes the result of a lexicalisation process or a conflation of two roots/stems. Note that grammatical morphemes containing a vowel only are quite common in Georgian. Monoconsonantal affixes are also quite common. On an even larger scale, the association of consonants with the lexical part of the morpheme and vowels with the grammatical part seems quite obvious (see the forms given in (2)).

3.2. The phonemic inventory

During the history of the Georgian language, the phonemic inventory has not undergone many changes. During the last ten centuries, the phoneme set was reduced by three elements. The semivowel /y/ was lost, the bilabial spirant /w/ merged with one of the variants of the labio-dental /v/ and the velar stop /q/ converged with the spirant /x/, although this sound is still retained in mountainous dialects of Georgian (Vogt 1961). The latter process caused the emergence of some homonymous forms.

(8)	*Modern Georgian*	*Old Georgian*	
	xeli	qeli	'hand'
	xeli	xeli	'stupid'
	xerxi	qerxi	'skill'
	xerxi	xerxi	'saw'

Since the Georgian vowel system is much simpler than the consonant system, the vowel system will be considered first.

3.2.1. The vowel system

The five vowels of Georgian can be characterised by two parameters: height and roundness.

(9) *Vowels*

Low	a	low, open, slightly fronted [a] (French *patte*)
High unrounded	i	front, spread lips, high, between close and half close [ɪ] (English *bit*)
High rounded	u	same height as /i/ [ʊ] (English *book* with marked lip rounding)
Mid unrounded	e	front, spread lips, between half close and half open [ɛ] (English *get*)
Mid rounded	o	same height as /e/ [ɔ] (German *Glocken*)

(Akhvlediani 1949)

Schematically, the Georgian vowel system can be presented as a triangle.

(10) *The vowel inventory*

i	[ɪ]			u	[ʊ]
e	[ɛ]			o	[ɔ]
		a	[a]		

Length is not a relevant feature of the vowel system and hence does not play a role either in the phonological processes or in accent assignment.[5] There are no diphthongs. There are some alternations involving vowels. Some of the alternations are phonologically conditioned, but most of them are morphologised, participating in ablaut in a verbal paradigm, e.g. the /e/ ~ /i/ alternation in pairs such as /da-grex-s/ 'he will twist' ~ /da-grix-a/ 'he twisted'. In this case, the opposition between future and past tenses, in addition to inflectional suffixes /-s/ vs. /-a/, is expressed by the vowel alternation /e/ ~ /i/. Among the vowel alternations, the following can be listed:

[5] Phonetic lengthening of vowels is observed before voiced fricatives and voiced stops, and especially before nasal sonorants (Akhvlediani 1949:282).

(11)			*NOM*	*GEN*	
	a)	o ~ v or ø	mindor-i	mindvr-is	‘field’
			xoxob-i	xoxb-is	‘pheasant’
	b)	a ~ ø	bal-i	bl-is	‘cherry’
	c)	e ~ ø or i	švel-i	švl-is	‘deer’
			c’avedi ‘I went’	c’avida ‘he went’	

The high vowels /i/ and /u/ are considered as more marked than the other vowels since they do not undergo any changes or deletion. Two distributional generalisations about the vowels /i/ and /u/ also substantiate this claim. They are given in (12).[6]

(12) a) Although the vowel /u/ is commonly found in the final position of monosyllabic lexical words, e.g. /bu/ ‘owl’, /ru/ ‘brook’, /χ’ru/ ‘deaf’, its occurrence in the final position of polysyllabic lexical words is very restricted.

b) There are no monosyllabic lexical words with final /i/. Although there are two words, /k’i/ ‘yes’ and /vin/ ‘who’, which seem to satisfy this condition, they are in fact not lexical words: the first is a confirmation particle and the second is a pronominal element. In monosyllabic words, the vowel /i/ is always a grammatical morpheme, e.g. /d-is/ (GEN) < /da/ ‘sister’, /č’r-i-s/ ‘somebody cuts’ < /č’ra/ ‘to cut’, etc.

These two observations suggest that the distribution of the high vowels is sensitive not only to the syllabic structure of a word (monosyllabic vs. disyllabic), but also to the asymmetry between lexical and grammatical morphemes. The vowel /u/ is preferred in lexical morphemes, while the vowel /i/ in grammatical morphemes. This morphologically motivated asymmetry between /i/ and /u/ is quite evident in grammatical affixes. The vowel /i/ is often found in inflectional affixes, while /u/ occurs only in derivational affixes, e.g. /u-/ is the particle of negation in forms such as /u-kud-o/ ‘without hat’, /u-namus-o/ ‘shameless’, and the objective version marker in forms such as /u-k’eteb-s/ ‘somebody does something to somebody else’.

There is a small set of monosyllabic, monomorphemic words in Georgian. All of them are vowel-final.[7] These words have a preference for final /u/, not /i/. This regularity is expected for acoustic and perceptual reasons. The vowel /u/ is more sonorous, and therefore more salient than /i/. This was confirmed by acoustic and perceptual studies on the Georgian vowels (Žgent’i 1956). In addition, phonetic studies on the long-distance influence of vowels on consonants reveal that the rounded vowels /o/, and especially /u/, have more influence on consonants than others; Georgian listeners could identify the trace of the deleted vowels in the cases when such vowels were labial, especially the vowel /u/ (K’iziria 1985). Thus, the

[6] The other three vowels, /a e o/, do undergo deletion and are characterised by quite free distribution throughout the word.

[7] In general, neither monosyllabic nor other types of non-derived words are consonant-final in Georgian.

occurrence of the vowel /u/ in prominent position, that is in final position of monomorphemic words, is more plausible than the occurrence of the vowel /i/ in the same position.

3.2.2. Hiatus

One basic phonotactic restriction on monomorphemic words is that two adjacent vowels are disallowed.[8] Two vowels can appear across morpheme boundaries. It is interesting that the prefix vs. suffix asymmetry shows up in such cases. Hiatus is tolerated at prefix # stem boundary, while at the stem # suffix boundary it is not. The most common way of resolving hiatus across morpheme boundaries is epenthesis. In most cases, the epenthetic segment is the sonorant /v/, e.g.

(13)	*NOM*		*Derived word*	
	rje	'milk'	me-rje-v-e	'milkman'
	t'χ'e	'forest'	me-t'χ'e-v-e	'forester'
	uto	'iron'	a-uto-v-eb-s	'somebody irons'

The epenthetic element, i.e. the sonorant /v/, is an interesting element in its own right, since its distribution, both diachronically and synchronically, can explain several facts of Georgian phonotactics. I will return to the behaviour of the sonorant /v/ when considering the consonant system, and especially the group of sonorants.

In a few cases, an epenthetic bilabial stop /b/ is found, e.g.

(14)	ezo	'yard'	m-ezo-b-el-i	'neighbour'

The insertion of /v/ is prohibited, however, when the following morpheme starts with either of the rounded vowels /o/ or /u/, e.g.

(15)	šina	'inside'	sa-šina-o	'home, domestic'

In general, all vowels except for /o/ and /u/ may be preceded or followed by any consonant: in monomorphemic words no */v/ + /o/ or */v/ + /u/ combinations are accepted. This restriction can be related to a general phonotactic restriction on adjacent bilabials in Georgian. To avoid combinations of */v/ + /o/ or */v/ + /u/ across morpheme boundaries, the sonorant /v/ is deleted.

(16)	tagv-i	'mouse'	sa-tag-ur-i	'mouse trap'
	tav-i	'head'	sa-ta-ur-i	'title'

Another strategy to avoid the hiatus across morpheme boundaries is vowel deletion.

[8] There are a few monomorphemic loan words with vowel complexes in Georgian, e.g. /musaipi/ 'talk' (Arabic), /paipuri/ 'porcelain' (Greek), /p'aik'i/ 'pawn' (Persian), /daira/ 'tambourine' (Arabic), /maudi/ 'cloth' (Arabic), etc. There are also cases with vowel complexes which are the result of compounding, e.g. /č'aobi/ < /č'a-obi/ 'swamp', /daisi/ 'sunset' < /da-isi/, etc. (see Uturgaidze 1976:40).

(17)

NOM	*GEN*	
jma	jm-is	‘brother’
xe	x-is	‘tree’

In general, there are no assimilation processes between vowels and vowel-consonant combinations. Literary Georgian does not allow assimilation processes, due to the strong influence of the orthographic normativity (Akhvlediani 1949). However, some assimilation processes are found in dialectal forms, e.g.

(18)

Literary Georgian	*Imeruli dialect*	
c’a-i-ɣ-o	c’eiɣo	‘he took something’
mo-i-t’an-a	meit’ana	‘he brought something’

3.2.3. Labials

There are some phonetic processes suggesting that the opposition involving labiality is an important one in Georgian phonotactics. Sequences of rounded vowels are disallowed both in monomorphemic environments and across morpheme boundaries. A dissimilation process occurs when two rounded vowels appear adjacently in derived words.

(19)

indo-uri	>	indauri	‘turkey’
sa-uto-o	>	sautao	‘iron desk’

The restriction on the occurrence of adjacent labials holds not only for vowel combinations, but also for vowel-consonant (see (41)) and consonant-consonant (see section 3.5) combinations. This observation suggests that there is a formal identity of labiality in consonants and roundedness in vowels.[9]

With regard to the observation on the labials, the Georgian data fit the cross-linguistic generalisation that is formulated as the Obligatory Contour Principle. That is, segments that are ‘identical’ are not permitted in a sequence (see Chapter 2).

3.2.4. Conclusions

The following generalisations summarise the paradigmatic and syntagmatic characteristics of the Georgian vowel system. There is very little assimilation between vowels and there are no diphthongs. Monomorphemically, adjacent vowels are also disallowed. The vowel alternations are either the result of morphological affixation or themselves have a morphological function.

The high vowels /i/ and /u/ have a defective distribution that is sensitive both to the number of syllables and the morphological composition of a word. They do not undergo the deletion process of Modern Georgian.

---

[9] In recent years a number of researchers have argued for a (partial) identification of Place features for consonants and vowels (see Reighard 1972, Campbell 1974, Anderson & Ewen 1987, Clements 1989, Levelt 1994, van de Weijer 1996).

The bilabial correlation is one of the most important, conditioning several phonological processes across morpheme boundaries.

3.3. The consonant system

As has already been pointed out, there are 28 consonants in the phonemic inventory of Georgian. For classification, it is convenient to divide them into three major sets: (i) stops and affricates, (ii) fricatives and (iii) sonorants.

Plosives: stops and affricates. The Georgian plosives can be characterised by place of articulation (according to which bilabial, dental, alveolar, palato-alveolar and velar plosives are distinguished) and laryngeal feature (distinguishing voiced, voiceless and glottalised sounds). The Georgian plosives are presented in (20).

(20) *Stops and affricates*

	Bilabial	Dental	Alveolar	Palato-alveolar	Velar
Voiced	b	d	j	ǰ	g
Voiceless	p	t	c	č	k
Glottalised	p’	t’	c’	č’	k’

From an acoustic point of view, the Georgian voiced obstruents are characterised by a low degree of voicing. This can be confirmed by transcriptions of Georgian speech by foreign listeners. The Georgian voiced obstruents are almost always perceived as either glottalised or voiceless (Žgent’i 1956). Some phonetic studies suggest that the voiceless consonants are aspirated (Žgent’i 1956). Thus, the classification according to laryngeal feature can be as follows: non-aspirated, aspirated and glottalised. More phonetic studies need to be carried out to establish which feature is more adequate for the description of Georgian consonants, [voice] or [aspiration].

Georgian has word-final devoicing. Since no minimal word ends in a consonant in Georgian, devoicing occurs in grammatical affixes, e.g. /v-a-k’et-eb/ is pronounced as [vak’etep] ‘I do something’, where /b/ belongs to the thematic suffix /-eb/ and /k’ac-ad/ is pronounced as [kac-at] ‘man ABL’, where /-ad/ is the Ablative case marker. Devoicing does not occur when the suffixes are followed by another suffix, for instance in forms such as /v-a-k’et-eb-di/ [v-a-k’et-eb-d-i] ‘I was doing’, and /k’ac-ad-a/ [k’ac-ad-a] ‘like a man’.

A few words must be said about the affricates, of which there are six in Georgian: /j c c’/ (alveolars) and /ǰ č č’/ (palato-alveolars). There are some cases when affricates appeared historically as a result of the merger of stops and fricatives, e.g.:

(21)

Old Georgian	*Modern Georgian*	
at-švidmet’i	čvidmet’i	‘seventeen’
at-samet’i	camet’i	‘thirteen’

In addition, there are cases of de-affrication.[10] Evidence for this comes from dialectal forms.

(22)	*Literary Georgian*	*Imeruli dialect*		
	sc'ori	st'ori	'right'	
	sc'avla	st'avla	'to study'	
	daǰdoma	daždoma	'to sit'	
	ocdaati	ozdaati	'thirty'	(Žgent'i 1956:237)

This is not to say that affricates represent mere combinations of stops and fricatives. Phonotactic restrictions discussed below indicate that affricates are single segments (see (23) and (24)). Experimental studies show that the phonetic duration of affricates is no longer than that of simple stops (Žgent'i 1956). Thus, Georgian affricates, like affricates of other languages, can be represented as segments linked to one timing slot or root node (see e.g. van de Weijer 1996).

Recently, Kehrein (1999) has proposed the elimination of the class of affricates from the phonological description and their inclusion in the natural group of stops. Even though, for the sake of convenience, the affricates are grouped together with stops in the above description, there are some phonotactic restrictions that suggest that affricates form a natural class of their own. For instance, the combination of a coronal stop followed by an affricate is not permitted in Georgian, e.g. the clusters *dc and *ǰd are ill-formed, while the reverse order is accepted, e.g. the clusters *cd* and *ǰd* are attested. There is another restriction in combination with fricatives. Affricates can be preceded by fricatives but never followed by them. For instance, the clusters *sc'* and *šǰ* are attested, while the clusters *c's and *ǰš are not. Both regularities also apply across a vowel. All these restrictions hold only within a lexical morpheme and can be violated across morpheme boundaries. The phonotactic restrictions concerning the combination of coronal and affricate consonants are depicted in (23) and those concerning the combination of fricative and affricate consonants are illustrated in (24).

(23) a) *coronal stop + (V) + [coronal stop + fricative]$_{\text{affricate}}$
b) [coronal stop + fricative]$_{\text{affricate}}$ + (V) + coronal stop

(24) a) *[coronal stop + fricative]$_{\text{affricate}}$ + (V) + fricative
b) fricative + (V) + [coronal stop + fricative]$_{\text{affricate}}$

In both (23) and (24), the OCP seems to be at work. In (23a), the restriction is imposed on adjacent coronal stops, and in (24a) the restriction holds for adjacent fricatives. Thus, optimal combinations are consonants with non-identical manner specification; in the case of stops, the restriction is even more specific, because it applies only to coronal stops. More such constraints, referring to place of articula-

[10] The claim that de-affrication in Georgian dialects is a later development is substantiated by Old Georgian data, where corresponding segments are affricates, as in Modern Georgian.

tion of stops, are considered in the following discussion of the syntagmatic regularities of consonants. These two observations (i.e. (23) and (24)) demonstrate the relevance of distinguishing affricates as a separate class (see also Butskhrikidze & van de Weijer 2001b).

Fricatives. Fricatives can be characterised by place of articulation (by distinguishing between alveolar, palato-alveolar and velar sounds) and laryngeal feature (by distinguishing between voiced and voiceless sounds). The fricatives are presented in (25).

(25) *Fricatives*

	Alveolar	Palato-alveolar	Velar
Voiced	z	ž	ɣ
Voiceless	s	š	x

There are two phonemes which have not yet been considered. One is the glottalised uvular fricative /χ'/, alternatively transcribed as a glottalised uvular stop (i.e. /q'/). This sound can be grouped together with the velar fricatives. It behaves like the fricative /x/ and its voiced counterpart, in that it can form a particular type of complex segments known as harmonic groups (see also below) when combined with anterior consonants. Another is the voiceless laryngeal /h/. It is the only laryngeal phoneme in Georgian. The distribution of /h/ is restricted to word-initial position, but even in this position it tends to delete in Modern Georgian. The following correspondences exemplify this:

(26)

Old Georgian	*Modern Georgian*	
hazri	azri	'mind'
hambavi	ambavi	'story'
hasak'i	asak'i	'age'
hambor	ambori	'hug'

Most of the words beginning with the consonant /h/ are loan words, e.g. from Greek: /hipoteza/ 'hypothesis', /harmonia/ 'harmony', etc. This sound never occurs in consonant sequences. Because of its defective distribution, /h/ will not be considered in the following discussion of consonant phonotactics.

Sonorants. There are 5 sonorants: the liquids /r l/, the nasals /n m/ and the labio-velar /v/. The sonorant /v/ has the allophones [v ɸ w], depending on the phonetic context:

(27) a) /v/ has the allophone [v] in word-initial position and between vowels, e.g. /viri/ [viri] ‘donkey’, /venaxi/ [venaxi] ‘wine-yard’, /tavi/ [tavi] ‘head’, /telavi/ [telavi] ‘the name of a town’.

b) /v/ has the allophone [ɸ] when followed by voiceless (aspirated or glottalised) consonants, e.g. /v-pikr-ob/ [ɸpikrop] ‘I think’, /v-c’er/ [ɸc’er] ‘I write’.

c) /v/ has the allophone [w] when preceded by an obstruent, especially by dorsal consonants, e.g. /χ’vavili/ [χ’wavili] ‘flower’, /kva/ [kwa] ‘stone’. The allophone [w] in such words is actually pronounced as a secondary articulation on the preceding consonants. Thus, more precise phonetic transcriptions of the words are [χ’^{w}avili] and [k^{w}a], respectively.

The characterisation of /v/ as a sonorant can be established on the basis of its distribution and behaviour in phonological processes. They are discussed in section 3.3.1.

In languages with long consonant sequences sonorants are often observed to be syllabic. Investigating the syllabic status of Georgian sonorants is an important issue for an adequate analysis of long consonant sequences. Consequently, the following section addresses the status of sonorants. The behaviour of the sonorant /v/ is considered extensively in a separate section. Two claims are made:

(28) a) Georgian sonorants are phonetically syllabic in consonant sequences. The claim is based on phonological processes, distributional patterns, and historical and comparative evidence. This especially concerns the most sonorous consonant /r/.

b) Combinations of C + /v/ can be treated as complex labialised segments, i.e. C^{w}. The claim is based on phonological, distributional and historical evidence. Reduplication data also substantiate this claim (see Chapter 5).

3.3.1. Syllabicity of sonorants

Akhvlediani (1949) assumes the phonetic syllabicity of the Modern Georgian sonorants. He argues that the syllabicity of the sonorants is context-dependent. It is realised only in clusters when the sonorant is surrounded by less sonorant consonants, e.g.

(29)

NOM	DAT		
naɣm-i	naɣm-s	[naɣms]	‘mine’
ipn-i	ipn-s	[ipn̩s]	‘ash tree’
saxl-i	saxl-s	[saxl̩s]	‘house’
tetr-i	tetr-s	[tetr̩s]	‘white’

The forms which demonstrate the syllabicity of the sonorants contain clusters in derived word-final position (forms in the Dative case). I argue that the syllabicity of

sonorants can also be assumed in consonant sequences occurring in initial position in monomorphemic words. Thus, the sonorant /l/ in /k'lde/ [k'l̩de] 'rock' is phonetically syllabic, as is the sonorant /r/ in /trtvili/ [tr̩tvili] 'hoar-frost'. To strengthen this claim, historical and comparative evidence as well as synchronic phonological processes and distributional regularities will be considered.

In their study of the correspondences between four languages of the Kartvelian language group and their investigation of the historical development of these languages, Gamkrelidze & Machavariani (1965) propose that Proto-Kartvelian sonorants are syllabic in certain phonetic contexts.[11] An explanation of the regular correspondences between the Kartvelian languages, ablaut in the verbal system and present-day syncope in the nominal paradigms in Modern Georgian is related to the syllabicity of sonorant consonants. Below I give some examples of Proto-Kartvelian syllabic sonorants, with correspondences in Kartvelian languages. The examples are taken from Gamkrelidze & Machavariani (1965).

(30)	*Proto-Kartv.*	*Georgian*	*Megrelian*	*Laz*	*Svan*	
	tr̩t	trt-ol-a	tirt-ol-i	tirt-ini		'trembling'
	k'l̩de	k'lde	k'irda/k'irde		k'oǰ	'rock'
	m̩ze	mze			miž/mež	'sun'

I will demonstrate that sonorants of Modern Georgian have retained their syllabicity in words like *trtola*, *k'lde* and *mze* (see e.g. (39) for *trtola* and (31) for *k'lde*).

Phonetically syllabic [r̩] and [l̩] are assumed to be present in the mountain dialects and some lowland dialects of Georgian (Gamkrelidze & Machavariani 1965). The following correspondences are considered to be illustrative:

(31)	*Old Georgian*	*Modern Georgian*	*Khevsuruli*	*Ingilouri*	
	grgoli	rgoli	girgoli		'ring'
	grk'ali	rk'ali	girk'ali	k'irk'ali	'arc'
		ɣrma	ɣirma		'deep'
		trtvili	tvirtvili	tirtul	'hoar-frost'
		črdili	čirdili		'shadow'
		k'lde	k'ilde		'rock'

The sonorants /r/ and /l/ seem to be syllabic in both Old Georgian and Modern Georgian. The corresponding Khevsuruli and Ingilouri forms with the vowel /i/ before the sonorants /r/ and /l/ substantiate this claim.

There are some phonological processes and distributional patterns which substantiate the claim about the syllabicity of Georgian sonorants, especially of the sonorant /r/. Consider in this respect the following: "The syllabicity of syllabic consonants never arises spontaneously from a marginal consonant, as far as I can ascertain. The source of the syllabicity is always a vowel. By far the commonest process

[11] Asatiani (1983) suggests that stops and fricatives are also syllabic in Proto-Kartvelian.

of origin is syllabic syncope, that is, loss of a vowel and shift of syllabicity to one of its margins" (Bell 1978:105).

Georgian has some phonological processes in which vowel loss is related to immediately following sonorants. One of these processes is syncope.[12] It occurs in the Modern Georgian nominal paradigm. Words with the stem-final syllables *-al*, *-ar*, *-el*, *-er*, *-an*, *-en*, *-am*, *-em*, *-ol*, *-or*, *-on*[13] undergo vowel deletion when followed by vowel-initial suffixes (case markers, the plural suffix, etc.). One of the requirements for syncope to occur is that a word must contain at least two syllables. Some examples of syncope in the nominal paradigm are given in (32).

(32)			
	NOM	mercxal-i[14]	'swallow'
	ERG	mercxal-ma	
	DAT	mercxal-s	
	GEN	mercxl-is	
	INST	mercxl-it	
	ABL	mercxl-ad	

As shown in (32), syncope occurs in three cases: the genitive, instrumental and ablative. In all of these cases suffixes are of the -VC type: /-is/, /-it/ and /-ad/, respectively.

Syncope does not occur when the stem does not end in a sonorant. Consider the case in ((33).

(33)			
	NOM	k'amat-i	'debate'
	ERG	k'amat-ma	
	DAT	k'amat-s	
	GEN	k'amat-is	
	INST	k'amat-it	
	ABL	k'amat-ad	

Another phonological process is truncation, which also occurs in the nominal paradigm. Vowels are lost in words ending with vowels other than /i/ (the nominative case marker) and /u/. In case words have two final syllables containing sonorants, both processes (truncation and syncope) apply simultaneously, e.g. /t'omara/ 'sack' in GEN /t'omr-is/, INST /t'omr-it/, ABL /t'omr-ad/. In these cases, two vowels are deleted, the stem-final /a/ and the stem-medial /a/.

---

[12] Syncope not only occurs in nominal paradigms (e.g. in the GEN, ABL and INST cases), but also with other types of suffixes with a -VC type structure, for example when the plural suffix /-eb/ is added to a stem, e.g. /bal-i /'cherry' ~ /bl-eb-i/ 'cherries'.

[13] There are some exceptional cases when syncope occurs in words with a stem-final obstruent /b/, e.g. /k'ak'ab-i/ 'partridge' ~ /k'ak'b-is/ (GEN) and /xoxob-i/ 'pheasant' ~ /xoxb-is/ (GEN).

[14] The nominative case marker /-i/ always follows consonant-final stems and never causes the reduction of stem-final vowels. In this respect, the nominative case marker is different from other vowel-initial suffixes.

In both the processes described above, a vowel deletes only when it is followed by a sonorant. The syncope process is also characteristic of verbal forms.

(34)	še-i-p'χ'ar-i	'arrest IMP'	še-i-p'χ'r-ob	'you will arrest'
	ga-č'er-i	'cut IMP'	ga-č'r-i	'you will cut'

The vowel-deletion process, which takes place under adjacency to sonorants, indicates the syllabicity of Georgian sonorants.

There are interesting distributional characteristics of the sonorants observed in clusters. The co-occurrence of the sonorants seems sensitive to the number of obstruents in a cluster. Two observations can be made about the co-occurrence of the sonorant /v/ with other sonorants, depending on the length and constituency of a cluster:

(35) a) In a cluster CS_1S_2 (obstruent + sonorant + sonorant), S_1 is always the sonorant /v/, while S_2 is either /n/, /l/ or /r/, e.g. /k'vn/ in /k'vnesa/ 'to moan', /k'vn/ in /kvnet'a/ 'to bite', /gvr/ in /gvrit'i/ 'turtle-dove', /c'vr/ in /c'vrili/ 'petty, small', /c'vl/ in /c'vlili/ 'mite', /xvr/ in /xvreli/ 'hole'. Note that in most of the cases obstruents are dorsals, which, in general, are the best landing sites for labialised consonants (Maddieson 1984). The generalisation is related to the claim that the sonorant /v/ is a secondary articulation on the preceding consonant. I return to this issue by considering extensive data in the following section.

b) In clusters with two or more obstruents (e.g. $C_1S_1C_2S_2$), the order of sonorants is different. The most sonorous sonorant /r/ takes the first position and the sonorant /v/ can only be the second sonorant of the cluster, e.g. /drt'v/ in /drt'vinva/ 'to grumble', /grgv/ in /grgvinva/ 'thunder' and /brč'χ'v/ in /brč'χ'viali/ 'sparkling'.

With regard to these generalisations, it is interesting to examine the combinations where C_1 is filled by a harmonic group or an /s/ + obstruent cluster. Both combinations count as one obstruent for these generalisations, e.g. /**px**vn/ in /pxvnili/ 'powder', /**sx**vl/ in /sxvla / 'chop off'. These patterns are in accordance with the generalisation formulated in (35a). Thus, these clusters could be formalised as CS_1S_2. I will discuss both types of clusters in more detail later.

From the case of sonorant distribution generalised in (35b), one could argue that a cluster of the type $C_1S_1C_2S_2$ cannot be analysed as one unit, but instead is divisible into sub-parts, as follows: C_1S_1/C_2S_2. The integrity of long sequences of the type $C_1S_1C_2S_2$ is challenged since the sonorant /r/ occupies the S_1 position and seems to be syllabic in this context. I will attempt a better formalisation of this assumption in Chapter 6, and for the time being, merely mention that for the examples given in (35b), the division would be as follows: *dr/t'vinva*, *gr/gvinva*, *br/c'χ'viali.*

Apart from the syllabicity of /r/, both generalisations (35a) and (35b) suggest the treatment of harmonic clusters and /s/ + obstruent clusters as one element C

(consonant). I return to these consonant sequences in the following sections and in Chapter 5. These generalisations also clearly illustrate that the sonorant /v/ occurs in consonant sequences predominantly after a dorsal consonant and can be treated as a secondary articulation on the dorsal consonant. The behaviour of the sonorant /v/ is discussed in the following section.

In addition, the sonorant /r/ displays interesting phonotactic behaviour in consonant sequences. For instance, sequences of the stop + fricative type are never separated by a sonorant, e.g. sequences such as **brz*, **pls*, **glz* are not attested, whereas sequences of the type fricative + stop are always separated by the sonorant /r/. Thus, sequences such *zrd* in /zrda/ 'to grow' and in /zrdiloba/ 'politeness' and *xrt'* in /xrt'ili/ 'gristle' are commonly found in Georgian (Nebieridze 1975). It seems that sequences violating the SSP, i.e. clusters of the fricative + stop type, are separated by syllabic /r/, e.g. /r/ in /xrt'ili/ is syllabic (see also the examples in (37)).

The behaviour of the sonorant /r/ is quite complex in Georgian and I will consider it in detail.

The sonorant /r/ can precede and follow any consonant or vowel. It is described as a 'fleeting' sonorant (Vogt 1961). In other words, it can appear optionally when it is surrounded by consonants with identical laryngeal specifications.

(36)	prta	~	pta	'wing'
	grdemli	~	gdemli	'anvil'
	brjeni	~	bjeni	'wise'
	brjaneba	~	bjaneba	'order'

As shown in the examples in (36), the sonorant /r/ can disappear when preceded and followed by consonants with identical laryngeal specifications.

'/r/-loss' does not take place when /r/ is preceded and followed by consonants which do not have the same laryngeal specification. Thus, the presence of the sonorant /r/ is obligatory in the following forms.

(37)	brč'χ'ali[15]	'claw'
	k'rjalva	'reverence'

In addition, the presence of the sonorant /r/ is obligatory when it is surrounded by identical consonants, e.g. in forms such as:

(38)	trtvili	'hoar-frost'
	grgvinva	'thunder'

According to Ertelishvili (1970), stems with the sonorant /r/ between identical consonants are always derived from reduplicated forms.

[15] '/r/-loss' is attested in dialectal forms, but only after assimilation has taken place: /brč'χ'ali/ > /p'rč'χ'ali/ > /p'č'χ'ali/ 'claw'.

(39) *Stems with /r/*

	Literary Georgian		*Dialectal form*	
Stems	C_irC_i		CVrCVC	
trt	trt-ol-a	‘to tremble’	tar-tal-eb-s	‘tremble’
	Literary Georgian		*Paradigmatically related form*	
prp	prp-en-a	‘to adore’	par-pat’i	‘to fly around’
			pr-en-a	‘to fly’

The examples given in (39) suggest that the sonorant /r/ in (38) is syllabic.

The phonological and phonotactic properties of the Georgian sonorants, especially those of the sonorant /r/, suggest that they should be considered as syllabic consonants. However, the syllabicity of Georgian sonorants has to be substantiated by phonetic studies.

3.3.2. The sonorant /v/ as a secondary articulation

Nepveu (1994) and Bush (1997), after observing that many of the clusters in Georgian contain /v/, and that it varies in pronunciation between [v], [ɸ] and [w], suggest that it is a ‘defective segment’. Nepveu argues that it is specified only for labial place of articulation, and acquires other features from the preceding consonant. It can be treated as dependent segment, i.e. a secondary articulation on the preceding consonant. I accept the analysis of the sonorant /v/ as a secondary articulation in consonant sequences and substantiate the claim with additional data from phonological processes, distributional regularities and historical evidence, and with reduplication data presented in Chapter 5.

Phonological processes involving the sonorant /v/ include metathesis, which is extensively discussed in Butskhrikidze & van de Weijer (2001a). Here I merely give a short introduction to the metathesis process relevant for the present discussion of the status of /v/. /v/, which is part of the thematic suffix in verb forms, violates integrity of the root when followed by the infinitival suffix /-a/. For example, the third person singular of the root /xar/ ‘to gnaw’ is /xr-av-s/ (with deletion of the root vowel). The infinitival form, however, is /xvr-a/ (with deletion of both the root vowel and the vowel in the thematic suffix), where the thematic suffix consonant /v/ occurs between the two root consonants.

(40) *Regular metathesis in verb forms*

ROOT	PRES 3rd SG (/-av/ THEM SUFF)	INF (/-a/ INF SUFF)	
xar	xr-av-s	xvr-a	‘to gnaw’
k’ar	k’r-av-s	k’vr-a	‘to tie’
xan	xn-av-s	xvn-a	‘to plough’
k’al	k’l-av-s	k’vl-a	‘to kill’
sxal	sxl-av-s	sxvl-a	‘to chop off’
jer	jr-av-s	jvr-a	‘to move’

In addition, there are some isolated forms displaying metathesis. There are cases when metathesis has occurred as a result of a diacronic change, e.g. /rva/ ‘eight’ > /at-rvamet’i/ ‘eighteen’ (Old Georgian) > /tvramet’i/ (Modern Georgian). There are cases of metathesis in dialectal forms too, e.g. /gač’irveba/ ‘difficulty’ (Literary Georgian) ~ /gač’ivreba/ (Imeruli dialect).

There are a few striking conditions on /v/-metathesis: the root must end in a sonorant other than /m/ (e.g. metathesis does not occur with a root like /xed/ ‘to see’) and the root should not start with a labial (metathesis does not occur with a root like /ber/ ‘to blow up’). The condition can be formalised as follows: C /r l n/ /v/ > C /v/ /r l n/. The condition is related to the constraint which bans two bilabials monomorphemically. I will return to this constraint later. It is interesting to observe that consonants preceding /v/ are predominantly dorsal (see (40)). Dorsal consonants are the best landing sites for labialisation cross-linguistically (Maddieson 1984).

Note that the metathesis process just described is closely related to the stem-vowel deletion process. One of the factors preventing vowel deletion in general in Georgian is avoidance of homonymous forms, e.g. in the form /k’erv-a/ ‘to sew’. Vowel deletion and consequently metathesis do not occur, because homonymous forms */k’vra/ ‘to sew’ and /k’vra/ ‘to bind’ would emerge. Thus, even though all conditions for the metathesis process are met in the form /k’erv-a/, it does not take place, because the language avoids creating homonymous forms.[16]

There are a number of other processes affecting the sound /v/, e.g. /v/-loss; /v/-alternation with /o/; /v/-epenthesis, etc. They are discussed briefly below.

/v/-loss: stem-final /v/ is lost when followed by affix beginning with the round vowels /o/ or /u/. Consider the paradigmatically related forms in (41).

(41)

NOM		*Derived word*		
tav-i	‘head’	meta-ur-i	(*me-tav-ur-i)	‘leader’
p’at’iv-i	‘honour’	p’at’i-osan-i	(*p’at’ivosan-i)	‘honest’

[16] The same constraint holds for the syncope process discussed in section 3.3.1. For instance, the genitive form of the noun /kari/ ‘wind’ is /kar-is/. Stem-vowel deletion is expected, but does not occur, because there already exists the form /kris/, which means ‘blows’. Thus, the stem vowel is retained in the form /kar-is/ to avoid homonymy.

‘/v/-loss’ under adjacency to labialised vowels can be related to ‘/v/-loss’ when followed by the bilabial nasal /m/ (e.g. /gamo-tkv-am-s/ ‘somebody is pronouncing’, but /gamo-tkma/ ‘pronunciation’), and to the distributional restriction on this sonorant with regard to the bilabial stops. The following discussion will demonstrate that combinations of bilabial consonants with /v/ are disallowed. It is interesting to note that, similarly to the case of hiatus, there is asymmetry between prefix # stem and stem # suffix contexts. While monomorphemic and stem # suffix environments do not tolerate two adjacent bilabials, the prefix # stem context allows such combinations, e.g. in form such as /v-mušaob/ ‘I am working’ and /v-mecadineob/ ‘I am studying’. Thus, the combination of *vm* is allowed across prefix # stem boundary; while in other contexts it is unettested.

/v/-alternation with /o/: the substitution of the vowel /o/ with /v/ occurs in derived words of the following type:

(42)	*NOM*	*GEN*	
	mindor-i	mindvr-is	‘field’
	p’amidor-i	p’amidvr-is	‘tomato’
	nior-i	nivr-is	‘garlic’
	nigoz-i	nigvz-is	‘nut’

The substitution of the /o/ by /v/ does not occur when the /o/ is preceded by labial consonant (recall that labial + /v/ is an ill-formed cluster). Hence one can find /sap’oni/ ‘soap’ ~ /sap’nis/, but not /*sap’vnis/. For the same reason, the substitution of the vowel /o/ by the sonorant /v/ is blocked in case it is followed by the bilabial sonorant /m/, e.g. in /diɣomi/ ‘toponym’ ~ /diɣmis/ (GEN).

/v/-epenthesis: to resolve hiatus, the sonorant /v/ is inserted between successive vowels, e.g. /ǰado/ ‘witchcraft’ ~ /a-ǰado-v-eb-s/ ‘somebody does witchcraft’.

There is some historical evidence for /v/ serving as a secondary articulation on consonants, especially on dorsals. According to some studies (e.g. Gamkrelidze & Machavariani 1965), there was a distinction between syllabic /u/ and non-syllabic /u̯/ in Proto-Kartvelian. The merger of the non-syllabic /u̯/ with the sonorant /v/ happened in a later period of the development of Georgian. The Modern Georgian sonorant /v/ corresponds to Old Georgian non-syllabic /u̯/. The correspondences are illustrated in the examples given in (43).

(43)	*Old Georgian*	*Modern Georgian*		
	sik’u̯dili	sik’vdili	[sik’^{w}dili]	‘death’
	eku̯si	ekvsi	[ek^{w}si]	‘six’
	varsk’u̯lavi	varsk’vlavi	[varsk’^{w}lavi]	‘star’
	čᵘ̯en	čven	[č^{w}en]	‘us’
	k’ᵘ̯ali	k’vali	[k’^{w}ali]	‘trace’

As shown in the examples in (43), just as in those in (40), [^{w}] appears after dorsal consonants in almost all cases, and it seems, is a cognate to the Old Georgian /u̯/.

The important consequence of establishing the status of /v/ as a secondary articulation in consonant sequences is that the combination of C + /v/ can be described not as a true cluster, but as a complex segment, i.e. a labialised consonant, the sequence of a consonant followed by [^{w}].

3.4. The minimal word

The definition of the minimal word is language-specific. An interesting property of the Georgian minimal word is its morphological constituency. To use the term *monomorphemic word* for Georgian is not especially insightful, because only vowel-final stems are monomorphemic, whereas consonant-final stems require a vowel (the nominative case marker *i*) to give a well-formed minimal word. Thus, in Georgian, words have two types of morphological structure: (i) word = stem, e.g. /da/ 'sister', /k'alata/ 'basket' and (ii) word = stem + the nominative case marker, e.g. /saxl-i/ 'house', /xel-i/ 'hand'. Morphological constituency plays an important role in the phonotactic organisation of a word. In the following sections I consider the word-initial and word-final positions in relation to the morphological and phonotactic structure of the word.

In many languages, lexical words of only one mora or syllable are avoided: a minimal bimoraic/disyllabic requirement is imposed (McCarthy & Prince 1986b, Kenstowicz 1994, among others). There are several types of evidence for such a disyllabic minimality constraint in Georgian.

Georgian has only vowel-final monosyllabic words of the CV, CCV, CCCV type, e.g. /xe/ 'tree', /rk'o/ 'acorn', /brge/ 'tall', etc.); there are no lexical words of the CVC type. Consonant-final stems always receive a suffixal vowel to form a well-formed minimal word, as in /k'ac-i/ 'man', /xel-i/ 'hand', etc. Vowel-final stems do not have a nominative case marker. Thus, they surface as bare stems, e.g. /magida/ 'table', /deda/ 'mother', /taro/ 'shelf', /sok'o/ 'mushroom'. The generalisation that words of the CVC type do not exist in Georgian indicates that the minimal Georgian word is of the CVCV type, i.e. disyllabic. This observation is interesting with regard to the maximum number of consonants in consonant sequences. As I argue later in this chapter, 'true clusters' in Georgian are maximally biconsonantal.

Since the minimal word is by definition (McCarthy & Prince 1986b) coextensive with a foot, another argument supporting the proposal of disyllabic minimality in Georgian could be stress assignment. According to Tevdoradze (1978), primary stress always falls on the first syllable of a word. In polysyllabic words, secondary stress occurs. In four-syllable words, secondary stress falls on the second syllable from the end, in five-syllable words it occurs on the third syllable from the end, and in six-syllable words on the fourth syllable from the end. According to this observation, a trochaic foot can be formed in polysyllabic words. That is to say, the constituent responsible for stress assignment in Georgian is a disyllabic trochaic foot. This could be formalised as follows:

(44)	a)	(σ́ σ)	déda	‘mother’
	b)	(σ́ σ) σ	déda-li	‘hen’
	c)	(σ́ σ)(σ̀ σ)	déda-èna	‘mother tongue’
	d)	(σ́ σ)(σ̀ σ)σ	déda-švìl-oba	‘motherly’
	e)	(σ́ σ)(σ̀ σ)(σ σ)	déda-búd-ian-ad	‘with all the family’

The stress patterns in (44) illustrate that words containing more than two syllables are morphologically complex. Monomorphemic trisyllabic and longer words of Georgian origin are very rare. These are mostly loan words, e.g. /ko’nst’it’ucia/ ‘constitution’, /p’arlament’i/ ‘parliament’, etc. This observation was confirmed by studies on the phonotactic structure of Georgian nominal and verbal stems (Ertelishvili 1970, 1980).

Additional evidence for the size of the minimal word comes from morphophonological processes. I will consider the formation of inalienable constructions in Georgian. A brief introduction to this matter is in order. “The prototypical morphological operation is affixation to a base. In most cases affixation occurs without regard to the phonological nature of the base. Of course, once affixation takes place, phonological rules can come into play. But in general, affixation occurs earlier and the phonology is left with the task of assigning a phonetic representation consistent with the rules and constraints of a language to the result” (Kenstowicz 1994:622).

However, there are cases in which the affixation process itself must take account of the phonology of the base. If the required structure does not obtain, affixation fails to occur. This is the case with the inalienable construction in Georgian.

Only kinship terms participate in inalienable possessive constructions. To form the construction, the terms should meet the following two conditions: they should be at least disyllabic, and they should have the vowel /a/ stem-finally. The latter is a general characteristic feature of Georgian kinship terms. The lexical items able to form the inalienable construction are as follows: /mama/ ‘father’, /deda/ ‘mother’, /bija/ ‘uncle’, /deida/ ‘aunt’, /mamida/ ‘sister of father’, /bicola/ ‘uncle’s wife’, /bebia/ ‘grandmother’, /babua/ ‘grandfather’. A sample paradigm of inalienable possessive constructions is given below.

(45)	*SG*		*PL*	
	mama-čemi	‘my father’	mama-čveni	‘our father’
	mama-šeni	‘your father’	mama-tkveni	‘your father’
	mama-misi	‘her/his father’	mati mama	‘their father’

The inalienable constructions are forms with the first person and second person singular and plural cases. An exception occurs in the third person plural, where /*mama-mati/ is ill-formed and the correct form is /mati mama/, with the pronoun preceding the noun and occurring separately. The same happens with other kinship terms that allow the inalienable possessive construction, such as /mati deda/ ‘their mother’, /mati bebia/ ‘their grandmother’, etc.

Monosyllabic kinship terms such as /da/ ‘sister’ and /jma/ ‘brother’ fail to allow the inalienable possessive construction. They take possessive pronouns independently and precede them, e.g. /čemi da/ ‘my sister’, /čemi jma/ ‘my brother’. Forms such as /*da-čemi/ and /*jma-čemi/ are ill-formed.

In one of the dialects of Georgian, Imeruli, where the same terms are disyllabic (/daia/ ‘sister’; /jamia/ ‘brother’), affixation applies and the constructions /daia-čemi/ and /jamia-čemi/ are well-formed (Butskhrikidze 1995). The blocking of the affixation process therefore seems to be conditioned by the disyllabic requirement of the base.

The process of monosyllabic lengthening provides further evidence for the disyllabicity of minimal words. In yes-no questions, whenever the last word of the question is monosyllabic, it is lengthened to two syllables by repeating the vowel. The process occurs if and only if the word in question is monosyllabic.

(46) *Monosyllabic lengthening*

a) k’art’opils pckvnis? ‘Is he peeling potatoes?’
[prckniis]

b) k’art’opili dač’ra? ‘Did he cut potatoes?’
[dač’ra]

In the example in (46a), lengthening of the monosyllabic word /pckvnis/ occurs, and it is pronounced with two *i*’s [prckniis], while in the example in (46b), lengthening does not occur, because the word /dač’ra/ is disyllabic. For more about the process, see Bush (1997). Finally, the reduplication process discussed in Chapter 5 also substantiates the claim that the minimal word is disyllabic in Georgian.

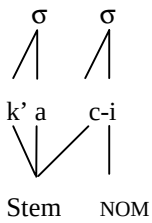
To conclude, the evidence discussed above suggests the existence of a disyllabic minimal word constraint in Georgian.

3.4.1. Word-final position

As has already been mentioned, a well-formed minimal word in Georgian must be vowel-final. The final vowel is either the nominative case marker or part of the stem. The occurrence of a single consonant or a consonant sequence is disallowed in word-final position. There are some exceptions in adverbs, most of which are sonorant-final (/c’in/ ‘in front’, /xval/ ‘tomorrow’, /gušin/ ‘yesterday’), or have voiceless obstruents (stops or fricatives) in final position (e.g. /zevit/ ‘up’, /kvevit/ ‘below’, /k’argad/ [k’argat] ‘well’). In the latter case, /-ad/ [at] is the ablative case marker and since we are not dealing with derived forms and their phonotactic patterns these cases will not be considered in the thesis.

As for the structural characteristics of the Georgian stem, the most important observation is that stems are predominantly consonant-final (Appendix 1 provides extensive data on the structural patterns of the nominal and verbal forms of Georgian). Syllable boundaries are not aligned with morphological boundaries in Georgian, e.g. /k’ac-i/ ‘man’ is syllabified as /k’a.ci/. The syllabic and morphological constituency of this word is depicted in (47):

(47) /k'aci/ 'man'



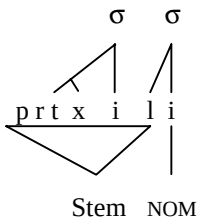
The schema in (47) is a representation of the mismatch between syllable and morphological boundaries. It also illustrates that consonantal material is within the stem domain. Notice that the relation that might hold between the consonants cannot be expressed in terms of syllables because the syllable boundary locates these consonants (i.e. the consonants /k'/ and /c/) in different constituents. In that case, the formulation of additional constraints operating on the adjacent constituents would be needed. The mismatch depicted in (47) has direct consequences for the possible types of analyses of the consonantal patterns of a language. Two options can be explored. One is an analysis which accounts for the study of consonantal patterns within the stem domain. The other is an analysis which accounts for the study of consonantal patterns in terms of a syllable constituent. The former is the one proposed and advocated in this thesis. The analysis is called the Gradual Consonant Analysis, and is extensively discussed in Chapter 6. The latter option has been considered in previous analyses of Georgian consonant clusters, and is summarised and discussed in Chapter 7. As for the stem, it can end in any of the 27 consonants (i.e. except /h/) and maximally form a five-member sequence. The patterns of stem-final consonant sequences are summarised in Appendix 3.

3.4.2. Word-initial position

All 33 phonemes of Georgian can appear in word-initial position. Words of Georgian origin are almost always consonant-initial. There are some words with a vowel in initial position, but most of them are loan words (mostly from Persian or Latin). Unlike word-final position, word-initial position is characterised by long consonant clustering, with sequences of up to six members.

Word-initial position corresponds to stem-initial position in Georgian (see section 3.5 for exceptional cases), while word-initial position does not correspond to syllable-initial position. The mismatch is illustrated by way of the word /prtxili/ 'careful' and is formalised in the schema in (48).

(48) /prtxili/ ‘careful’



The schema in (48) shows that in /prtxili/ the first two consonants, /p/ and /r/, cannot be syllabified together with /t/ and /x/, since the syllable-initial position (onset) is already filled by two consonants (/t/ and /x/).[17] Thus, the word cannot be exhaustively syllabified. We are left with two word-initial consonants (/p/ and /r/) which are not part of the syllable. Nevertheless, the whole *prtxil* sequence is part of a structural constituent stem that encompasses the four-member cluster as a whole. The following questions arise with respect to consonant sequences of this type:

(49) a) Are there constraints on consonant combinations in such long consonant sequences, or is their constituency random?
b) If the co-occurrence is not random, then what are the principles governing them?

These are questions I will address in the remainder of the thesis. However, for the present, it is important to keep in mind the domain mismatches depicted in (47) and (48).

3.4.3. Conclusions

To conclude, two important observations have been made with respect to the characteristics of the Georgian minimal word.

The disyllabicity of the minimal word is substantiated by phonological (accent assignment), phonetic (monosyllabic lengthening) and morphological (the alienable/inalienable construction) evidence. Additional evidence will come from the reduplication process discussed in Chapter 5. The Georgian minimal word can be formally presented as a general disyllabic template of the $C_1V_1C_2V_2$ type.

Syllable-final position coincides with word-final position, while syllable-initial position does not coincide with word-initial position. The mismatch can be related to the fact that the morphological boundary is not aligned with the syllable boundary in Georgian. Consequently, the phonotactics of Georgian can be better interpreted not only by looking at syllable constituency, but by also taking into account the mor-

[17] At this point, the optional occurrence of the sonorant /r/ and the treatment of harmonic clusters (here *cx*) as complex segments are disregarded. These will be considered later, in the analysis of consonant sequences (see Chapter 6).

phological constituency of a word. Stem-initial position in Georgian corresponds with word-initial position, while stem-final position does not correspond with word-final position (except vowel-final stems). This is exactly the opposite situation to the one described above, i.e. the syllable/word correlation. In the following section, I discuss word-initial, i.e. coinciding with stem-initial, consonant combinations.

3.5. General introduction to consonant syntagmatics

Apart from the study of consonant grouping that was carried out in section 3.3, another classification referring to the place of articulation is needed to account for the syntagmatic patterns of consonant combinations. Thus, we arrive at the table in (50), in which primary focus is on place and manner of articulation.

(50) *Consonants grouped according to place of articulation and manner*

Manner / Place	Stop			Nasal	Fricative					Liquid	
Bilabial	b	p	p’	m							
Alveolar	d	t	t’	n	j	ǰ	c’	z	s	l	r
Palato-alveolar					c	č	č’	ž	š		
Velar	g	k	k’					ɣ	x		
Uvular									χ’		

There are many co-occurrence restrictions of the OCP type in Georgian, referring primarily to place of articulation.

Geminates are disallowed in monomorphemic contexts.[18] The classification discussed below is based on Uturgaidze (1976). It accounts for the majority of the Georgian data.

The first restriction concerns the labial sounds (including the labio-dental sonorant /v/). This class is denoted as C1. It contains the sounds /b p p’ m v/.

It is not permissible to have two adjacent labial consonants in a monomorphemic environment. Any combination between the members of this class is disallowed. Thus, *mb-, *mp-, *mp’-, *mv-, *bm-, *bp-, *bp’-, *bv-, *pm-, *pb-, *pv-, *vm- (recall that /v/ is lost when followed by /m/), etc. are not found. Recall that combinations of rounded vowels, as well as rounded vowels and the sonorant /v/, are also disallowed (see sections 3.2.2 and 3.2.3).

The second class, C2, contains the coronals, /d t t’ j c c’ z s ǰ č č’ ž š/, in which three places of articulation are distinguished:

[18] There are no geminates in Georgian. However, identical consonants do occur across morpheme boundaries, e.g. /xaz-ze/ ‘on the line’, /v-varǰišob/ ‘I am training’, /mat-tan/ ‘with them’. In order to avoid gemination, one of the consonants undergoes deletion, e.g. /t’ani-samosi/ /*t’anis-samosi/ ‘cloth’, etc. Foreign words with geminates always enter the Georgian lexicon with only one consonant, e.g. /alegoria/ ‘allegory’, /k’lasi/ ‘class’, etc.

Dentals: /d t t'/
Alveolars: /j c c' z s/
Palato-alveolars: /ǰ č č' ž š/

C2 has two restrictions on the co-occurrence of its members:

a) Combinations of homorganic phonemes are not accepted (i.e. the same restriction as for labials);

b) A posterior coronal may precede an anterior coronal, but never follow it. Thus, the combinations *št, ǰd, cd* are attested, while *tš *dǰ, *dc never occur.

The third class, C3, contains the sonorants, /r l n/. It has the following restriction: the members of the class never combine (i.e. the same restriction as for the first and the second classes).

The velar and uvular consonants are grouped together. This class is referred to as C4, and contains /g k k' ɣ x χ' h/. It has the following restriction: the members of the class never combine with each other (i.e. the same restriction as for the first, the second and the third classes).

To conclude, the generalisation is as follows:

(51) Obstruents with identical place of articulation never combine.[19]

Two additional factors are of importance when consonant sequences are discussed.

(52) a) The places of articulation of the members of the consonant sequence.
b) The laryngeal specification.

With regard to the first factor, regressive and non-regressive consonant sequences can be distinguished. 'Regressive' refers to anterior + posterior consonant sequences (i.e. labial-coronal, coronal-dorsal, labial-dorsal). 'Non-regressive' refers to posterior + anterior consonant sequences (i.e. dorsal-coronal, coronal-labial, dorsal-labial). With regard to the second factor (i.e. (52b)), clusters are defined as 'homogeneous' or 'heterogeneous'. In 'homogeneous' clusters, members share the laryngeal feature. Thus, they are voiced, voiceless or glottalised. Most scholars (Akhvlediani 1949, Vogt 1961, Melikishvili 1997) consider regressivity and homogeneity to be the preferred patterns for Georgian consonant sequences.

Both conditions, regressivity and homogeneity, are met in the consonant combinations referred to as 'harmonic clusters'. Because of the long tradition of the study of the harmonic clusters in Georgian, and because of the importance of their representation in my analysis, the following section is devoted to these.

[19] In other words, sequences of homorganic consonants are disallowed in Georgian. The claim does not concern the class of sonorants since combinations of e.g. coronal obstruents and coronal sonorants are well-formed. For instance, sequences such as *dn, dr, t'l*, etc. are attested.

3.5.1. Harmonic clusters

3.5.1.0. Introduction

The history of the study of the harmonic clusters in the Georgian linguistic literature goes back to the nineteenth century. Over this period of time, harmonic clusters have been defined in many different ways, but the two requirements, regressivity and homogeneity, are recognised in almost all proposals (Khundadze 1901, Marr 1925, Akhvlediani 1949, Machavariani 1965 and Uturgaidze 1971, among others). Clusters of [−dorsal] [+dorsal] obstruents are called harmonic in Georgian because they share a laryngeal specification. There are two types of harmonic clusters, referred to as types A and B. They differ in their constituency. Type A refers to combinations of stops, affricates and fricatives with the velar stops /g k k'/ and type B refers to combinations of stops, affricates and fricatives with the fricatives /ɣ x χ'/. In some studies (Marr 1925, Machavariani 1965), combinations of fricatives with dorsal obstruents are also treated as harmonic clusters. In these clusters, the regressive order is respected and, taking into account the fact that laryngeal feature is less distinctive for fricative consonants than for stops, homogeneity is maintained. I incorporate these additions and present the clusters as follows:

(53)

Type A			*Type B*		
bg	pk	p'k'	bɣ	px	p'χ'
dg	tk	t'k'	dɣ	tx	t'χ'
jg	ck	c'k'	jɣ	cx	c'χ'
ǰg	čk	č'k'	ǰɣ	čx	č'χ'
zg[20]	sk		zɣ	sx	
žg	šk		žɣ	šx	

In order to define the status of the harmonic clusters, two types of evidence could be considered: phonological and phonetic. Phonological evidence is considered in this chapter. Phonetic evidence, which consists of a perceptual experiment on harmonic clusters, is discussed extensively in Chapter 4.

3.5.1.1. Phonological behaviour

There are several types of evidence: phonological processes, distributional facts, historical considerations, etc. that suggest an analysis of harmonic clusters as complex segments. The observations are considered below.

Harmonic clusters never occur across morpheme boundaries. They always belong to one morpheme, specifically to the lexical morpheme, the stem.

[20] I. Melikishvili has pointed out to me that harmonic clusters such as *zg* and *žg* are almost unattested. They were probably included in the set in order to mentain the symmetry with the Type B set.

(54) a) *Stem-initial position*

pkvil-i	‘flour’
t’k’bil-i	‘sweet’
dɣe	‘day’
čxir-i	‘stick’
cxvir-i	‘nose’

b) *Stem-final position*

ortkl-i	‘steam’
marc’χ’v-i	‘strawberry’
čončx-i	‘skeleton’
otx-i	‘four’
vepxv-i	‘tiger’

Harmonic clusters are the only obstruent sequences to appear in stem-final position. No other type of obstruent sequence is attested in this position (Vogt 1961, Deprez 1988).

Phonological processes affect both members of a harmonic cluster, i.e. in case one consonant of a harmonic cluster change, the other member also changes, such that the processes shown below, for instance, are not attested (Dzidzishvili 1966).

(55) *dg ↗ tg / ↘ dk *tk ↗ dk / ↘ tg

Thus, both members of a cluster undergo a change, e.g. the alternation *dg ~ tk* is attested in Old Georgian form such as *c’ar**dg**ra ~ c’ar**tk**ra* ‘introduced’ (Dzidzishvili 1966:232).

Some forms containing harmonic clusters have parallel forms with another type of harmonic cluster, e.g. both *bur**d**ɣuni ~ bur**t’χ’**uni* ‘muttering’ and *b**d**ɣvriali ~ bč’χ’vriali* ‘glitter, sparkle’ are attested in Modern Georgian.

Consonant sequences other than harmonic clusters are characterised by optional /r/-insertion (Vogt 1958, Chikobava 1971, Deprez 1988).

Harmonic clusters always syllabify together intervocalically, e.g. /si.t’χ’va/ ‘word’, /ce.cxli/ ‘fire’, while other obstruent clusters are never tautosyllabic, e.g. /mar.t’i/ ‘March’, /bev.ri/ ‘much’ (Akhvlediani 1949 and Žgent’i 1956, among others). Evidence for these syllabification patterns mostly comes from native speaker intuitions.

In reduplicated forms, harmonic clusters retain their complexity, e.g. /čkar-čkara/ ‘quickly’, /cxel-cxeli/ ‘hot’. Other types of clusters do not usually participate in reduplication. Reduplication is discussed in more detail in Chapter 5.

Harmonic clusters are found in all Kartvelian languages (i.e. Svan, Megrelian and Laz). There are many examples of correspondences of harmonic clusters between these languages.

(56)	*Georgian*	*Megrelian*	*Laz*	*Svan*	
	mat'χ'l-i	mont'χ'or	mont'k'or		'wool'
	txra	txorua	ontxoru	lištxri	'to dig'

As shown in the examples in (56), the harmonic clusters do not undergo any simplification processes, and have direct correspondences in all Kartvelian languages.

3.5.1.2. Conclusions

To conclude, several types of evidence: phonological processes, distributional facts and historical considerations suggest that harmonic clusters can be analysed as complex segments. Phonetic evidence is discussed in Chapter 4. The harmonic clusters are considered as complex segments in numerous studies, e.g. Machavariani & Gamkrelidze (1965), Deprez (1988), Bush (1997), Nepveu (1994), Cho & King (1997), van Lit (1988), among others. The harmonic clusters are treated as being parallel to the C + /v/ combinations by Gamkrelidze & Machavariani (1965). In the following sections, I strengthen this position with additional arguments from the patterns of long consonant sequences and phonotactic characteristics of the stems of the CVC type.

3.5.2. Consonant combinations in adjacency

It has already been mentioned that consonant sequences appear word-initially. There are no word-final consonant sequences. Since word-medial clusters show the subpatterns of word-initial clusters, I do not consider them in my analysis. For the patterns of stem-final word-medial consonant sequences, see Appendix 3.

In this section I would like to demonstrate that the complexity of word-initial consonant sequences is quite illusory in Georgian. The two-consonant clusters that seem to be the building blocks of much longer consonant sequences are of the type obstruent + sonorant. Thus they largely obey the SSP. The most frequently realised clusters are combinations of obstruent + sonorant, harmonic clusters and consonants that share a laryngeal specification and have [front] + [back] place of articulation. The clusters that do not obey the SSP are largely morphologically complex.

Since the domain of my analysis is the minimal word, two types of consonant sequences, /v/ + C and /m/ + C, will not be considered. In both cases, /v/ and /m/ are of morphological origin.

In monomorphemic words, /v/ can never be the first member of a cluster. It appears in word-initial position in clusters, but always as a prefix, e.g. /v-rčeb-i/ 'I am staying'. Two words, /vseba/ 'to fill' and /vrceli/ 'wide' are exceptions to this generalisation. Unlike /m/, the sonorant /v/ can appear in other positions in clusters, but in almost all cases /v/ has a morphological origin and is subject to the metathesis process. For instance, in /k'vra/ 'to bind', /v/ is derived from the thematic /-av/ morpheme, /k'r-av-s/ 'somebody binds somebody else' (see (40)).

The sonorant /m/ is a nominaliser and can precede any consonant. The morphological origin of /m-/ is obvious in forms such as /m-c'eral-i/ 'writer', /m-k'v'lel-i/ 'killer' and /m-c'vrtnel-i/ 'trainer'. However there are some words where it is difficult to trace the morphological origin of /m-/, e.g. in forms such as /mz-e/ 'sun', /mgel-i/ 'wolf' and /mt'red-i/ 'pigeon'.

Comparative data could shed some light on the constituency of the forms with /m/ + C clusters. Consider, for example, correspondences between forms of Modern Georgian and another Kartvelian language, Megrelian.

(57)	*Modern Georgian*	*Megrelian*	
	mxari	xuǰi	'shoulder'
	msxali	sxuli	'pear'
	mxali	xuli	'name of a meal'
	mč'adi	č'k'idi	'maize-bread'

The Megrelian forms lack the word-initial /m-/, which may suggest that the /m-/ in Modern Georgian is a grammatical morpheme with no direct correspondences in other Kartvelian languages.

In addition, there are correspondences between Literary Georgian and other Georgian dialects (e.g. the West Georgian dialects Acharuli, Guruli and Imeruli) involving parallel forms with and without word-initial /m-/. Consider the correspondences in (58). The data come from Gudava (1979).

(58)	a)	*Literary Georgian*	*Acharuli*	
		mze	ze	'sun'
		mc'are	c're	'hot, bitter'
		mk'lavi	k'lavi	'arm'
		mc'χ'emsi	c'χ'esi	'shepherd'
	b)	*Literary Georgian*	*Imeruli*	
		mta	ta	'mountain'
		mgeli	geli	'wolf'
		mdidari	didari	'rich'
		mšobeli	šobeli	'parent'

The dialectal forms systematically lack /m-/. These correspondences once again substantiate the claim that /m-/ should be a prefixal morpheme rather than a part of the lexical morpheme in Literary Georgian. Thus, sequences of the /m/ + C and /v/ + C type are not considered to be genuine clusters in Georgian.

Typological studies on consonant sequences suggest the following generalisation: "… all languages exhibit the following property: if clusters of *n* Cs are possible syllable-initially, then clusters of *n–1* Cs are also possible syllable-initially, and if clusters of *n* Cs are possible syllable-finally, then clusters of *n–1* Cs are also possible finally" (Blevins 1995:217; see also Greenberg 1978). Maximally, six-consonant

clusters can appear word-initially in Georgian. This implies that there are also two-, three-, four- and five-member clusters.

Studying the constituency of two-member to six-member consonant sequences reveals a quite transparent structure (see (59)).[21] Here I merely give a sample of the sequences beginning with the consonant /b/. I will try to generalise on the basis of these examples and point out interesting patterns.

(59) *Consonant sequences beginning with /b/*

		CC	CCC	CCCC	CCCCC
a)	Stop + stop	bg	–	**bdɣv**	bdɣvn
		–	–	–	bdɣvr
b)	Stop + affricate	bǰ	bǰɣ		
		bč’			
				bc’k’	
c)	Stop + fricative	bɣ	bɣv		
		–	bɣl		
		bz			
		bž			
d)	Stop + sonorant	bn			
		br	brg		
			brk’		
			brj		
			brt’		
			brm		
		–	–	**brč’χ’**	brč’χ’v

The consonant sequences given in bold are special in the sense that they do not have a shorter counterpart. There are three such sequences: ***bc’k’***, ***bdɣv***, ***brč’χ’***. For example, the *bc’k’* sequence occurs in word-initial position, while **bc’* is unattested. The same can be said about the other two cases: sequences such as **bd* and **brč’* are not attested word-initially. All these consonant sequences contain harmonic clusters (i.e. *c’k’*, *dɣ* and *č’χ’*, respectively). This indicates that *c’k’*, *dɣ* and *cˇχ’* appear as single units. This observation is an additional argument supporting the treatment of the harmonic clusters as complex segments. Hence, the representation of the sequences should be as follows: *c͡’k’*, *d͡ɣ*, *č͡’χ’*. After taking into account what we have said about harmonic clusters, we can analyse the *bc͡’k’* sequence as a two-member consonant sequence. In the case of *bdɣv*, another assumption about C + /v/ clusters is needed in addition to the assumption about harmonic clusters. I have already argued that the C + /v/ combinations are complex segments. Thus, even in the case of

[21] For extensive data see Appendix 2.

an apparent four-member sequence, we end up with a two-member one, i.e. *bdɣ͡ʷ*. Finally, consider the four-member *brč'χ'* sequence. The sequence is attested in only one word, *brč'χ'ali* 'claw', which has a parallel form *p'rč'χ'ali*, with optional 'fleeting' /r/ (recall the description of the sonorant /r/ as 'fleeting' in section 3.3.1). The sequence *p'cʶχ'* exists in parallel with the sequence *p'rč'χ'*. Thus, in these cases, as in the previous cases (i.e. as in cases of the *bc'k'* and *bdɣ͡ʷ* clusters), we are left with the two-member sequence *p'č'χ'*.

The same transparency of long consonant sequences is observed in the following generalisation, suggested by Gvinadze (1970).

(60) I /b p p' m/
II /r/
III /d t t' j c c' ǰ č č' z s ž š/
IV /g k k' ɣ x χ'/
V /v/
VI /r l m n/

One consonant from each set can combine in the strict order given and form maximally a six-member cluster, e.g. /brdɣvna/ 'to fight', /prckvna/ 'to peel'. Any set can be skipped, but the order between the sets should be respected.

Consonants in the sequence must have the same laryngeal specification and must be regressive. For instance, the five-member sequences are: *bdɣvr-*, *p'rcʶχ'v-*, etc.

As shown in (60), the sonorant /m/ is included in the first set. It seems that Gvinadze (1970) did not exclude word-initial /m/ + C combinations from her analysis. As I have mentioned, I do not consider /m/ + C combinations here, because of the morphological origin of /m/. The double occurrence of /r/, in both the second and sixth sets, is explainable by the optionality of the first /r/ in long consonant sequences.

It can also be seen from (60) that the first set consists of bilabial stops. This is remarkable, since we have already seen a number of distributional restrictions on bilabial consonants and rounded vowels in Georgian. The appearance of bilabials in the initial position of long consonant sequences can be related to the phonotactic restrictions that hold in general for the bilabials (including rounded vowels) in Georgian phonotactics. The initial position in the sequence is taken by bilabials because they have less combinatory power than other segments, such as coronals. The behaviour of the Georgian bilabials shows their marked character, and it is in accordance with the cross-linguistic observation that bilabials are marked obstruents (Melikishvili 1976).

The second consonant in a consonant sequence is the sonorant /r/, which is optional. For instance, it is optional in all six-member sequences. In general, in long consonant sequences it is optional when preceded and followed by consonants with an identical laryngeal specification. It is obligatory when preceded and followed by

non-homogenous obstruents. As was argued earlier, in such contexts the /r/ is syllabic.

The consonants from the third and fourth sets usually form harmonic clusters. They can be followed by the sonorant /v/ from set 5, which, as I have argued earlier, can be analysed as a secondary articulation on the preceding consonant. The longest consonant sequences are closed by a sonorant consonant from set 6.

Since the longest clusters of Georgian, as suggested in the discussion of the examples in (59), can be analysed as two-member clusters, the main topic of further discussion will concern two-member clusters. The following generalisation justifies the focus on two-member clusters:

(61) There are no three-member clusters which do not contain a sonorant.

Since Georgian sonorants are assumed to be syllabic in my analysis, apparent three-member sequences including a sonorant will be analysed as spurious sequences with the syllabic sonorant. For instance, the sequences *kld* in *klde* 'rock' and *prt* in *prta* 'wing' include the syllabic sonorants *l̩* and *r̩*, and in the analysis will be broken up by an unspecified vowel (see Chapter 6). Three-member obstruent sequences, e.g. *t'k'b*, *p'c'k'*, *bjg*, are considered as two-member sequences because they contain harmonic clusters *t'k'*, *c'k'*, *jg*, respectively.

Surface stem/word-initial two-member sequences are given in (62). The table depicts two characteristics of biconsonantal sequences discussed above, regressivity and homogeneity. The table does not include /m/ + C and /v/ + C sequences.

(62) *Possible (+) and impossible (–) initial biconsonantal clusters*

C_1/C_2	b	p	p’	d	t	t’	g	k	k’	j	c	c’	ǰ	č	č’	z	s	ž	š	ɣ	x	χ’	v	l	r	m	n
b	–	–	–	–	–	–	+	–	–	–	–	–	+	–	+	+	–	+	–	+	–	–	–	+	+	–	+
p	–	–	–	–	+	–	–	+	–	–	–	–	–	+	–	–	+	–	+	–	+	–	–	+	+	–	+
p’	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+	+	–	+
d	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	+	+	+	+	+
t	+	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+	+	+	+	–
t’	+	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	+	–	–
g	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	+	+	+	+	+
k	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	+	–	+	–	–	–	+	+	+	+	+
k’	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	+	+
j	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	+	+	+	+	+
c	+	–	–	+	+	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+	–	+	+	+
c’	+	–	–	–	+	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	+	+	+
ǰ	–	–	–	+	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	+	+	–	+	–
č	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+	+	+	+	–
č’	–	–	–	+	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	+	+	+
z	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	+	–	+	+	+
s	–	+	+	–	–	+	–	+	+	–	–	+	–	–	–	–	–	–	–	–	+	–	+	+	+	+	–
ž	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	+	+	+	–	–
š	–	+	+	–	+	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+	+	+	+	+
ɣ	–	–	–	+	–	–	–	–	–	+	–	–	+	–	–	–	–	–	–	–	–	–	+	+	+	+	+
x	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+	–	–	–	+	+	+	+	+
χ’	+	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	+	+
l	+	–	+	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	–	–	+	–
r	+	–	–	–	+	+	+	+	+	+	+	+	+	+	–	–	–	–	–	+	+	+	+	–	–	–	–
n	–	–	–	+	+	–	+	–	–	–	–	–	+		–	–	–	–	–	–	–	–	–	–	–	–	–

The shaded areas in some parts of the table denote high concentrations of pluses, which are the most frequently realised clusters. There are three such areas in the table: (i) the last five columns, (ii) two rows, (iii) the rest of the shaded areas. I will discuss each of these in turn.

(i) Most ‘+’s in (62) are concentrated in the last five columns of the table. These are combinations of consonants with the sonorants /v l r m n/, i.e. sequences of the C + sonorant type. These show that stem/word-initial clusters have a preference for rising sonority. As argued in section 3.3.2, sequences of C + /v/ can be analysed as complex, labialised segments.

(ii) Two shaded rows present clusters with *s* and *r* as their first member. These are clusters with falling sonority. The two are different in that the sequences of the /s/ + C type can be analysed as complex segments, while in /r/ + C sequences the sonorant /r/ appears to be syllabic and its occurrence as a first member of two-member sequences is related to a number of factors to be discussed below.

The frequency of the /s/ + C clusters can be related to the large number of old and recent borrowings from the Indo-European languages, e.g. Greek, Latin, English, etc. For instance, *sp* is attested in /spero/ 'sphere'; *sp'* is attested in /sp'ort'i/ 'sport' and /sp'irali/ 'spiral'; etc. Interestingly, the sequences of the /s/ + obstruent type behave as complex segments in Modern Georgian. Some of them are treated as harmonic clusters, e.g. *sk* and *sx* (see (53)). Reduplication patterns also provide additional evidence for treating these sequences as complex segments (see Chapter 5).

The sequences of the /r/ + C type are very common in Georgian. Melikishvili (1997) proposes an explanation for the appearance of the sonorant /r/ in stem/word-initial position. The arguments are briefly summarised below.

a) The first factor causing the appearance of /r/ in word-initial position is the transition from closed syllable to open syllable in the development of Georgian.

b) The second factor conditioning the appearance of /r/ in word-initial position is related to the process whereby the fricative /s/ in word-initial position in Old Georgian[22] changed into the sonorant /r/ in Modern Georgian.

(63)

Old Georgian	*Modern Georgian*	
sje	rje	'milk'
sjali	rjali	'groom'
sǰuli	rǰuli	'religion'
stveli	rtveli	'vintage'

In all of these examples the /s/ seems to have a morphological origin but the change to /r/ still needs to be explained.

c) Melikishvili (1997) relates the occurrence of /r/ cluster-initially to a cluster-simplification process. She says that the Proto-Kartvelian syllabic /ɣ/ became non-syllabic, which was followed by deletion of a preceding consonant, i.e. *****CɣC > CrC > rC:

(64)

Proto-Kartvelian		*Old Georgian*		*Modern Georgian*	
gɣk'ali	>	grk'ali	>	rk'ali	'arc'
gɣc'χ'ili	>	grc'χ'ili	>	rc'χ'ili	'flea'
k'ɣk'o	>	k'rk'o	>	rk'o	'acorn'
gɣgoli	>	grgoli	>	rgoli	'ring'

(Melikishvili 1997:33)

In section 3.3.1 I argue that in both cases ((63) and (64)) /r/ in Modern Georgian is phonetically syllabic.

(iii) The rest of the shaded areas refer to harmonic clusters.

Let us now consider some of the biconsonantal sequences which are not within the shaded areas of the table but are attested in Modern Georgian, such as *gz*, *cd*, *ct*, *tb*, etc. Ertelishvili (1970) and Uturgaidze (1976) convincingly argue that such sequences are derived from CVC stems. Consider the following forms:

[22] /s-/ in Old Georgian could have a morphological origin, as a prefix.

(65)	gz	gza	‘way’	gezi	‘direction, way’
	ct/cd	cda	‘to try’	e-cad-e	‘try you’
	tb/t’p	tbili	‘warm’	gan-t’ep-i-t	‘get warm (Old Georgian)’

These types of consonant sequences seem marginal to the Georgian language. Sometimes, as shown in the examples in (65), it is possible to trace the vowel between the consonants by looking at parallel forms of Modern or Old Georgian, but sometimes it is difficult to find such evidence, for instance in case of clusters such as *χ’b* and *k’b*. Comparative data shed some light on the origin of such sequences, e.g. the sequence *χ’b* attested in the word /χ’ba/ ‘jaw’ is reconstructed in Proto-Kartvelian as /χ’ab/; the corresponding form in Svan is /χ’ab/ (Sardschweladse & Fähnrich 1990). The consonant sequence *k’b*, which is attested in the word /k’bili/ ‘tooth’, can also be reconstructed as /k’Vb/, because the corresponding form in Megrelian and Laz is /k’ib-ir-i/. Thus, the form with a vowel between the consonants /k’/and /b/ occurs in Kartvelian languages.

To conclude, three types of biconsonantal sequences are commonly attested.

(66) *Preferred biconsonantal sequences*
- a) Obstruent + sonorant
- b) Harmonic clusters
- c) /s/ + obstruent
- d) /r/ + obstruent

Of these four types of consonant sequences, only the sequence of obstruent + sonorant (i.e. (66a)) is a genuine biconsonantal cluster, since, as advocated in this thesis, harmonic clusters are treated as complex segments. The sequence of /s/ + obstruent type is also considered as complex segment (see Chapter 5 for additional evidence) and /r/ + obstruent sequences are not considered as genuine biconsonantal clusters because of their derived nature.

3.5.3. Conclusions

The study of word-initial consonant sequences has shown that the complexity of consonant sequences is quite illusory in Georgian. The two-consonant clusters which seem to be the building blocks of much longer consonant sequences are of the type obstruent + sonorant. Thus they largely obey the SSP.

The most frequently realised clusters are combinations of obstruent + sonorant, harmonic clusters and consonants that share a laryngeal specification and have [front] + [back] place of articulation. Clusters that do not obey the SSP are largely of morphological origin.

3.6. Testing the hypothesis

In Chapter 2, I proposed that all two-member consonant clusters are derived by vowel deletion; their co-occurrence restrictions are derivable from stems of the CVC type. [23] The proposal was formulated as a hypothesis, which is repeated as (67).

(67) *Hypothesis*
If a language has clusters of the C_iC_j type, then the language will have stems of the C_iVC_j type.

The merits of this hypothesis are that it is easily refutable, easily testable and has predictive power. The following discussion is devoted to the verification of the hypothesis.

I have already formulated the basic characteristics of biconsonantal sequences in Georgian. Now I turn to consonant combinations across a vowel.

Firstly, I summarise the studies by Kobalava (1967) on Modern Georgian stems of the CVC type and Melikishvili (1997) on the typological study of the Kartvelian root structure. Secondly, I present the results of my findings. Finally, the restrictions on combinations of consonants at a distance and in adjacency will be compared.

3.6.1. Studies by Kobalava (1967)

The investigation of Kobalava (1967) was based on consonant combinations in stems of the CVC type. Using an explanatory dictionary of Georgian she studied all nominal and verbal stems of the CVC type. Particular attention was paid to consonants.[24] Of a potential 729 possible consonant combinations only 457 are attested (i.e. 63 %). The unattested 272 possibilities reflect the structural properties of Georgian phonotactics. These findings can be summarised as follows:[25]

a) Consonants from the same set do not co-occur in CVC stems (excluding combinations of identical consonants):

(68) *Stops and Affricates*
/b p p'/
/d t t'/
/j c c'/
/ǰ č č'/
/g k k'/

[23] Consonant clusters here refer to genuine clusters, and not ones which are the result of complex segment formation or morphological merger, i.e. the result of a lexicalisation process.
[24] Kobalava (1965) mentions a possible correlation between constraints on adjacent and distant consonant combinations, but does not investigate this further.
[25] There are some exceptions (mostly loan words) to the generalisations in (68) and (69), but they certainly reflect the basic phonotactic properties of Georgian stems.

Fricatives
/z s/
/ž š/
/ɣ x χ'/

b) Affricates and fricatives do not combine with each other: /j c c' z s ǰ č č' ž š/.

c) Some consonant combinations in a certain order are allowed, while the reverse order is unacceptable, e.g. /b p p'/ + V + /m v/ is attested, while */m v/ + V + /b p p'/ is unattested.

d) If the first consonant is a glottalised stop or an affricate, the second voiced consonant must have a place of articulation that is further to the front than that of the first consonant (restrictions for the second voiceless or glottalised consonants are not attested). The generalisation is expressed in (69).

(69)

p'	V	*b	*d	*j	*ǰ	*g
t'	V	b	*d	*j	*ǰ	*g
c'	V	b	d	*j	*ǰ	*g
č'	V	b	d	j	*ǰ	*g
k'	V	b	d	j	ǰ	*g

Thus, the combinations *t'b*, *c'b*, *k'b* are attested, while the combinations **t'p'*, **c'j*, **c'g*, etc. are disallowed.

e) The occurrence of two /r/'s or two /v/'s is disallowed within a CVC stem.

All possible (denoted by the symbol +) and impossible (denoted by the symbol –) consonant combinations within the stems of the CVC type are summarised in (70). The shaded areas denote the high concentrations of pluses. Note that most pluses appear in columns where the second consonant is a sonorant. Thus, in Georgian the realisation of the CVS (sonorant) structure is almost 100%. Recall that adjacent consonants also predominantly have a CS structure.

(70) *Consonant combinations within the stem of the CVC type (Kobalava 1967)*

C_2/C_1	b	g	d	v	z	t	k’	l	m	n	p’	ž	r	s	t’	p	k	ɣ	χ’	š	č	c	j	c’	č’	x	ǰ
b	+	+	+	+	+		+	+	+	+	–	+	+	+	+	–	+	+	+	+	+	+	+	+	+	+	+
g	+	+	+	+	+	+	–	+	+	+		+	+	–		+	–		–	+		+			+	+	+
d	+	+	+	+	+	–	+	+	+	+		+	+	+	–	+	+	+	+	+		+				+	
v	–	+	+	–	+			+	–	+	–	+	+	+		–				+		+				+	+
z		+	+	+	+	+	+	+	+	+	+	–	+	–	+	+	+			–	–	–	–	–	–	+	–
t	+	+	–	+		+	+	+	+	+			+	+	–	+		+		+						+	+
k’	+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–	+	–	+		+		+	+	+	
l	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+
m	–	+		+		+		+	+	+	–	+	+	+	+	–	+			+					+	+	
n	+		+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+				+	+	+	
p’	–	–	–	+			+	+		+	+	+	+	+	+	–				+			–	+		+	
ž				+	–		+	+	+	+	+	+	+	–	+					–	–	–	–	–	–		–
r	+	+	+	+	+		+	+	+			+			+	+	+		+	+		+		+	+	+	+
s	+			+	–	+	+	+	+	+	+	–	+		+	+	+	+		–	–	–	–	–	–	+	–
t’	+	–	–	+	+	–	+	+	+	+	+		+	+	+	+				+	+	+	–			+	–
p	–	+		+		+	+	+	+	+	–	+	+	+	+		+		+	+	+	+		+	+	+	
k	+	–	+	+	+	+	–	+	+	+			+	+	+	+	+		–	+	+	+			+	+	+
ɣ	+		+	+	+		+	+	+	+	+	+	+		+	+		+	–	+		+			+	–	+
χ’	+	–	+	+	+	+	–	+	+	+	+	+	+		+	+	–	–	+	+	+				+	–	
š	+		+	+	–	+	+	+	+	+	+	–	+	–	+		+			+	–	–	–	–	–	+	–
č		+	+	+	–	+	+	+	+	+		–	+	–	+	+	+		+	–	+	–	–	–	–	+	–
c			+	+	–	+	+	+	+	+		–	+	–	+	+				–	–	+	–	–	–	+	–
j	+	+		+	–		+	+	+	+		–	+	–		+		+		–	–	–		–	–	+	–
c’		–	+	+	–	+	+	+	+	+	+	–	+	–	+		+			–	–	–	–	+		+	–
č’		–	+	+	–	+	+	+	+	+	+	–	+	–	+		+	+		–	–	–	–	–	+	+	–
x			+	+	+	+	+	+	+	+	+		+	+	+	+		–	–	+	+	+		+	+	+	
ǰ	+	+	+	–	–	+	+	+	+	+	+	–	+	–	+	+	+	+	+	–	–	–	–	–	–	+	+

3.6.2. Studies by Melikishvili (1997)

Melikishvili’s studies on Kartvelian root patterns are of great interest for my analysis, not only because of the data and generalisations (in addition to those already observed by Gamkrelidze & Machavariani 1965, Fänrich 1978, etc.), but most of all because of her theoretical standpoint. When looking at the root/stem and the syllable structures of different types of genetically distinct languages, she establishes the following correlation: the basic phonotactic constraints for the stem/root and syllable structures are the same cross-linguistically.[26] While accepting the general idea, I will extend it and offer some additional proposals to strengthen it.

[26] The same type of claim is made on the basis of studies of the Semitic languages (see, e.g. Belova 1991, Diakonoff 1970, 1988).

The consequence of the correlation between the root/stem and syllable structures is to expect the Sonority Sequencing Principle – originally formulated for the syllable domain – also to function on the stem domain. Melikishvili demonstrates this for the Kartvelian CVC stems. Firstly, she discusses the distributional characteristics of sonorants in CVC stems. These occur in initial and final position in accordance with the general sonority schema: /m/ → /n/ → /l/ → /r/.[27] The sonorant /r/ is the most dispreferred in the initial position,[28] although it is often found in final position. The reverse situation holds with the least sonorous consonant /m/, which is more likely to occur stem-initially than stem-finally. These observations for Kartvelian stems are supported by similar studies on stem patterns of other languages (Indo-European (Magnusson 1967, 1979), Semitic (Greenberg 1970) and German (Twaddell 1939, 1941)).

More original are Melikishvili's generalisations on the distributional characteristics of stops in CVC sequences. Assuming the importance of defining the sonority scale not only for manner, but also for place of articulation features, she proposes the following constraint for accounting for the phonotactic restrictions within the stems of the CVC type.

(71) *The Compensatory Principle (CP)*
The co-occurrence patterns of heterorganic consonants across a vowel are such that if the first consonant is more sonorous than the second according to the laryngeal specification, the second consonant will be more sonorous according to the place of articulation. If the first consonant is more sonorous than the second according to the place of articulation, then the second consonant will be more sonorous according to the laryngeal specification (Melikishvili 1997:57, my translation).

The implications of the principle are illustrated with respect to the Kartvelian roots of the CVC type in (72):

(72)

Laryngeal Specification		Place of Articulation	
		Rising Sonority Regressive	Falling Sonority Non-regressive
Rising Sonority	Glottalised-voiced	p'-g	k'-b
	Voiceless-voiced	p-g	k-b
	Glottalised-voiceless	p'-k	k'-p
Falling Sonority	Voiced-glottalised	b-k'	*g-p'
	Voiced-voiceless	b-k	*g-p
	Voiceless- glottalised	p-k'	*k-p'

[27] Nebieridze (1974) proposes this sonority hierarchy.
[28] The absence of the sonorant /r/ in stem-initial position (in terms of syllable constituency, onset position) is attested in a number of languages, e.g. Basque, Armenian, Greek, Khunzic and Andic.

The combinations of k'-b, k-b and k'-p, as well as their reverse order b-k', b-k and p-k' are attested, while combinations violating the requirement do not occur. Combinations of *g-p', *g-p and *k-p' are not permitted. In the latter combinations, the sonority falls, according to both place of articulation and the laryngeal specifications, while in the former cases sonority increases in one of the two dimensions: it either increases in terms of place of articulation or in terms of laryngeal specification.

The CP holds for Modern Georgian consonant combinations across a vowel. As Melikishvili points out, the generalisation made by Kobalava follows from this principle. For ease of exposition I repeat the generalisation.

(73)

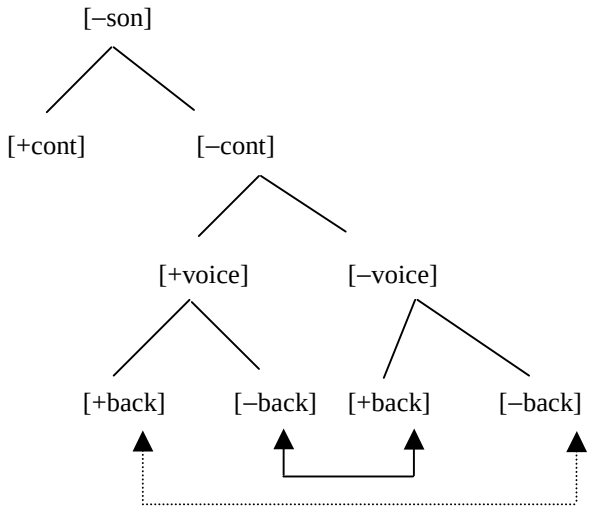
p'	V	*b	*d	*j	*ǰ	*g
t'	V	b	*d	*j	*ǰ	*g
c'	V	b	d	*j	*ǰ	*g
č'	V	b	d	j	*ǰ	*g
k'	V	b	d	j	ǰ	*g

Thus, **p'b*, **p'd*, **p'j*, **p'ǰ*, **p'g*, **t'd*, **dj*, etc. are not attested clusters; while *t'b*, *c'b*, *c'd*, etc. are well-formed clusters.

Melikishvili's proposal is justified not only by language data, but also from an articulatory-perceptual perspective. Two requirements desirable for any language structure, ease of perception and ease of articulation, are achieved in consonant combinations with different place of articulation and different laryngeal specifications (see dissociation of like consonants (Krupa 1967)).[29] Consonants that are too dissimilar are dispreferred in the same way as consonants that are too similar, e.g. a [−voice, −back] consonant does not combine with a [+voice, +back] consonant. This could be due to the fact that they are too distant. The balance is achieved when features are specified in alternating way, e.g. a [−voice, +back] consonant combines well with a [+voice, −back] consonant. The Compensatory Principle that captures the optimal combinations of consonants is depicted in (74). A dotted line denotes the dispreferred consonant combinations, while a solid line denotes the preferred ones.

[29] Zubkova (1990) establishes the existence of a tendency: combinations of vowels are characterised by harmony (homogeneity), while combinations of consonants are characterised by contrast (heterogeneity). This holds not only for adjacent consonants, but also for consonants at a distance.

(74)



Melikishvili's proposal is comparable to the hypothesis proposed by Saporta (1955).

> … the average frequency of a consonant cluster is a function of the difference between the phonemes in the cluster: low frequencies are expected for clusters which are either extremely similar or extremely dissimilar; high frequencies are expected for clusters which are at neither extreme (Saporta 1955:25).

The reasoning behind this is that the relative frequency of consonant clusters reveals a tendency on the part of any language system to produce speech in such a way as to consider the effort of both the speaker and the listener, the encoder and the decoder. Saporta says:

> In any consonant cluster, the situation of least effort for the speaker is that in which the successive phonemes are not similar; but this requires maximum effort on the part of the listener, who is then forced to make a series of fine discriminations. For the listener, the optimal situation is that in which the phonemes of a cluster differ as much as possible; but this requires maximum effort on the part of the speaker (Saporta 1955:25).

The Compensatory Principle balances combination of feature classes in consonant sequences and disallows too similar or too dissimilar consonant combinations. Thus, the CP can be regarded as one of the instantiations of the Balancing Principle, just like the SSP and the OCP, discussed in Chapter 2. The CP operates in Georgian and the Kartvelian languages (as advocated by Melikishvili 1997), but the cross-linguistic validity of this principle remains to be verified.

3.6.3. My study

In Chapter 2, I proposed the following hypothesis:

(75) *Hypothesis*
If a language has C_iC_j clusters, then the language will have stems of the C_ivC_j type.

As predicted, the restrictions within clusters are more constrained than between consonants across a vowel; compare the two tables given in (62) and (70), for example.

In order to test the hypothesis in (75), I used a dictionary containing 60,000 words (Chikobava 1986). I extracted all stems of the $C_1V_1C_2$ type; afterwards all word-initial biconsonantal sequences were extracted. The co-occurrence patterns in biconsonantal clusters were compared with the co-occurrence patterns across a vowel within $C_1V_1C_2$ stems.

The findings of the comparison were that there are no biconsonantal clusters whose members are not found within a CVC-type stem domain,[30] for example, Georgian does not permit stems such as **dVp'*, **tV j*, **t'aj*, **dVt'*, and, consequently, **dp'*, **tj*, **t'j*, **dt'* clusters are not allowed either. Notable exceptions are harmonic clusters, e.g. clusters such as *p'k'*, *p'χ'* and *t'χ'* are attested while stems of the type **pV'k'*, **p'Vχ'* and **t'Vχ'* are not. This observation provides further justification for the treatment of the harmonic clusters as complex segments.

Having established generalisations on consonant combinations in adjacency and across a vowel, I can spell out some restrictions that hold for consonant combinations both in adjacency and across a vowel. Members of the following phonemic sets do not combine with each other:

(76)	a)	Stops	/b p p'/ /d t t'/ /g k k'/
	b)	Affricates and fricatives	/j c c' z s ǰ č č' ž š/
	c)	Fricatives	/ɣ x χ'/

The restrictions concerning fricatives and dental stops are as follows: a sequence of dental stop + fricative is disallowed, while the combination fricative + dental is permitted. One restriction concerns bilabial consonants (including /v/): /b p p'/ + /m v/ is attested, but /m v/ + /b p p'/ is disallowed. Thus, the constraints on C_iC_j in both types of structures, C_iC_j (cluster) and C_iVC_j, are similar.

3.6.4. Conclusions

Similarities have been found between the composition of stems and consonant sequences, which suggests the hypothesis in (75).

[30] As I have said, clusters seem to be the result of vowel deletion. This holds only for true, i.e. core, clusters. Some 'clusters' like harmonic clusters, sequences of C + /v/ and /s/ + obstruent are treated as single segments.

In order to test the hypothesis, I compared the consonant co-occurrence restrictions in adjacency and across a vowel. The co-occurrence patterns in both contexts are quite similar, e.g. both show a preference for rising sonority, i.e. show the effect of the SSP. In addition, both types of restrictions have similar OCP effects.

The plausibility of the hypothesis has been confirmed on the basis of data from Georgian. In order to demonstrate the cross-linguistic validity of this hypothesis, it would be necessary to test it on a large corpus of genetically distinct languages.

The hypothesis leads to another sub-hypothesis:

(77) All consonant sequences (maximally biconsonantal) are derived, i.e. the result of vowel deletion.

I return to this claim in Chapter 6, which focuses on the analysis of the Georgian consonant sequences. This claim concerns only genuine biconsonantal clusters; as illustrated in this chapter, surface consonant sequences may be the result either of morpheme concatenation or of complex cluster formation, as is the case with harmonic clusters or with C + /v/ combinations. It is also possible that the language adopts morphologically complex loan words as monomorphemic ones. However such words would represent a restricted subset of the vocabulary and should be treated separately by taking into account the patterns of both host and donor languages.

3.7. General conclusions

The most salient observations about the Georgian word-level phonotactics are as follows.

Georgian, as a Grammatical language (i.e. a language with features of inflectional and agglutinative morphology), demonstrates the discrepancy between lexical and grammatical morphemes in terms of the use of the phonological inventory.

One way to explain the surface consonant sequences of Georgian is to carefully study the morphological structure of words where the sequences could be the result either of the deletion of a stem vowel (which generally happens when a vowel-initial affix is added to a root) or of mere addition of a consonantal affix to a root.

The disyllabicity of the Georgian minimal word is substantiated by phonological (accent assignment), phonetic (monosyllabic lengthening) and morphological (alienable/inalienable constructions) evidence. Additional evidence comes from the reduplication process discussed in Chapter 5. The Georgian minimal word can be formally presented as a general disyllabic template of the $C_1V_1C_2V_2$ type.

Georgian sonorants are syllabic in consonant sequences. This claim is based on phonological, ditributional, historical and comparative evidence. This especially concerns the most sonorant consonant /r/. A comprehensive phonetic study should be carried out to test this claim.

Combinations of C + /v/ can be treated as complex labialised segments. The claim is based on phonological, distributional and historical evidence. This claim is substantiated by reduplication data in Chapter 5.

Phonological processes, distributional facts and historical considerations suggest that harmonic clusters can be analysed as complex segments. Phonetic evidence discussed in Chapter 4 strengthens this claim.

The study of word-initial consonant sequences has shown that the complexity of consonant sequences is illusory in Georgian. Two-consonant clusters, which are the building blocks of much longer consonant sequences, are of the type obstruent + sonorant. Thus, they largely obey the SSP.

The most commonly attested clusters are combinations of obstruent + sonorant, harmonic clusters and consonants that share the laryngeal specification and are regressive ([front] + [back] place of articulation). Sequences not obeying the SSP are largely of morphological origin.

Similarities are found between the composition of stems and consonant sequences. These observations lead to the hypothesis that if a language has C_iC_j clusters, then the language will have stems of the C_iVC_j type.

The plausibility of the hypothesis has been confirmed on the basis of the Georgian data. The hypothesis entails that all consonant sequences (maximally biconsonantal ones) are derived, being the result of vowel deletion. I return to this claim in Chapter 6.

4 Phonetic Evidence for Harmonic Clusters: A Perceptual Experiment

4.0. Introduction

Harmonic clusters in Georgian consist of two obstruents, the first of which is a coronal or a labial and the second of which is a dorsal.[1] The obstruents share all voicing characteristics (see Chapter 3).

(1)

Type A (C + stop)			Type B (C + fricative)		
[+voi]	[–voi]	[glott]	[+voi]	[–voi]	[glott]
bg	pk	p'k'	bɣ	px	p'χ'
dg	tk	t'k'	dɣ	tx	t'χ'
jg	ck	c'k'	jɣ	cx	c'χ'
ǰg	čk	č'k'	ǰɣ	čx	č'χ'
zg	sk		zɣ	sx	
žg	šk		žɣ	šx	

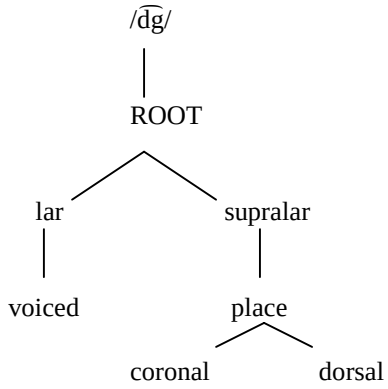
In the investigation of the special behaviour of Georgian harmonic clusters, several phonological arguments have been proposed to support their analysis as complex segments.[2] Some of these were considered in Chapter 3, section 3.5.1.

In addition to phonological evidence, phonetic claims about simultaneity of closure and release in harmonic clusters have been put forward by Kutelia (1956), Žgent'i (1965) and Aronson (1982). Both types of evidence led Deprez (1988) to the conclusion that harmonic clusters should be treated as complex segments, represented by a single root node, as shown in (2).

[1] This chapter is a revised version of Butskhrikidze & van Heuven (2001).

[2] Harmonic clusters are considered as complex segments in numerous studies (Gamkrelidze & Machavariani 1965, Bush 1997 and Nepveu 1994, among others).

(2)



Chitoran (1998) argues that if harmonic clusters are complex segments, their articulatory gestures should overlap temporally. As a result, the release burst of the first member of the cluster is suppressed by the occlusion phase of the second member of the pair, so that only the release burst of the second stop should surface. Also, due to the (partial) temporal overlap of the two obstruents in the harmonic cluster, the duration of the complex segment should be shorter than that of the corresponding (heterosyllabic) sequence of consonants.

Thus, in order to find acoustic evidence that would motivate the distinction between harmonic clusters behaving as single segments and not as simple sequences of consonants, Chitoran (1998) considered two parameters: the frequency of the occurrence of release bursts of the first and the second member of the sequence (Burst 1 and Burst 2) and duration. As tokens, near-minimal pairs containing word-initial or word-medial harmonic clusters and sequences of the same segments formed across a word boundary were considered.

(3)

Word-medial		*Sequences*	
datkma	'to agree'	albat kari	'probably wind'
ǰibgiri	'hooligan'	egeb gip'ovis	'perhaps he will find you'

The results of the experiment, frequency of occurrence of release bursts and durational measurements are given in (4) and (5), respectively.

(4) *Release burst*

	Word-initial	*Word-medial*	*Sequences*
Burst 1	95%	99%	94%
Burst 2	100%	93%	100%

(5) *Durational measurements*

	Word-medial	*Sequence*
dg, bg	166ms	140ms
tx, cx	192ms	196ms

Unexpectedly, there was no significant difference in frequency of occurrence of release bursts of the first and second stops in the two types of clusters. In fact, bursts were realised in more than 90% of the cases, regardless of the type of cluster. In addition, the duration of the harmonic clusters was not shorter than that of the corresponding C # C sequences. Moreover, if there were significant duration differences, they were in the unexpected direction, e.g. as shown in the measurements in (5), voiced harmonic clusters are longer than the sequences of the same segments across a word boundary.

Thus, the results of Chitoran's (1998) acoustic study show no significant differences for harmonic clusters vs. sequences, either in terms of the occurrence of release bursts or duration. She concludes that the acoustic study does not support the phonological treatment of harmonic clusters as complex segments, but, instead, indicates that harmonic clusters are sequences of segments.

It should be pointed out that merely counting the presence or absence of release bursts is a very crude analytical tool. In previous studies it has been found that C1 bursts in the complex segment were always present but had shorter durations and/or weaker intensities than their counterparts in heterosyllabic sequences (Butskhrikidze 1998b). This 'incomplete suppression' of the C1 burst is hard to reconcile with the notion of gestural overlap, but does argue for a different status of harmonic groups as opposed to C + C sequences. Also, the phonological environments in which the harmonic clusters and their corresponding C + C sequences were examined were not held constant. As shown in the examples in (3), the contexts in which harmonic clusters and consonant sequences are embedded do not contain the same number of syllables, and, in general, do not have similar phonetic environments. It is difficult to maintain phonetic sameness when tokens are actual words of a language. This type of problem usually affects the accuracy of measurements, and this factor might well have affected Chitoran's (1998) results. Therefore, the issue of the phonetic status of harmonic clusters remains unsettled.

The perceptual experiment described in this chapter is an attempt to test the phonological claim that harmonic clusters should be treated as complex segments.

4.1. An alternative approach

The goal of the experiment is to show whether harmonic clusters are perceptually cued differently from other types of consonant sequences, but similarly to simplex segments. The task of the experiment was to detect a CVC target embedded in three different types of non-word contexts: harmonic clusters (CV.CcV...), unrelated sequences (CVC.CV...), and single consonants (CV.CV...).[3] Two dependent variables were used to evaluate the experimental results: detection rate (hit rate in %) and detection latency (reaction time in ms).

[3] The dot between V and C represents a phonetic syllable boundary. Cc denotes a harmonic cluster.

If harmonic clusters are indeed complex segments, it should be difficult to detect a CVC target embedded within a CV.CcV… context. This is because the final C in the target string occurs in the onset of the next syllable and is (arguably) part of a single phonological segment (i.e. a harmonic cluster). Consequently, in terms of the two parameters, hit rate and detection latency, hit rates should be poor, while reaction times (for correct decisions) should be high.

When the target is embedded in an unrelated consonant string context (CVC.C…) detection of the CVC target should be easy (high hit rates and fast latencies) if the final C of the target is in the coda position of the test string, i.e. if the stimulus string has the structure CVC.C….

An intermediate degree of difficulty is predicted for the detection of the CVC target in a stimulus string where the final C of the target appears as a single consonant in the onset of the next syllable. The detection is relatively difficult because the listener has to resyllabify the final C, but less difficult than in the case of a harmonic cluster, as no decomposition of the cluster is required.

For instance, the target *k'ab* should be maximally easy to detect in the stimulus string *k'ab.ra.na* (final *b* is in coda position), most difficult to detect in *k'a.bga.na* (final *b* is part of a complex cluster in the onset of the following syllable) and intermediate in *k'a.ba.na* (final *b* is a single C in the next onset).[4]

4.1.1. Method

4.1.1.1. Stimuli

Targets were always sequences of the CVC type, embedded in different phonetic contexts. 32 sets (corresponding to the number of harmonic clusters; see (1)) were designed, each containing six non-words.[5] Three non-words in each set were target bearing, as follows:

(6)
a) One containing a harmonic cluster (harmonic Cc).
b) One containing the consonant string (unrelated CC), the first segment of which was identical with the beginning of the harmonic cluster.
c) One containing just the first segment of the harmonic cluster (single C).
d) The other three non-words in the set did not contain the target.

For instance, in the case of the target *tab* the set of six non-words was *ta.bga.mi – tab.ra.mi – ta.ba.mi – ta.ra.mi – da.ra.mi – ča.ra.mi*. The full set of non-words is included in Appendix 4.

192 non-words (32 sets x 6 non-words) were recorded on Digital Audio Tape (DAT) by a female native speaker of Standard Georgian (the author) in a soundproofed recording booth, using a Sennheiser MKH416 unidirectional condenser microphone. The recordings were then transferred to computer disk, downsampled to

[4] The syllabification patterns are based on native speaker intuitions and are rather tentative.
[5] Non-words were chosen for several reasons: to maintain sameness of the phonetic environment of a target, and to avoid morphological or word-frequency effects.

16 kHz, edited by means of a high-resolution waveform editor (*Praat* software; Boersma & Weenink 1996) and stored on disk.

4.1.1.2. Listeners

24 native Georgian listeners (15 female and 9 male) participated in the experiment. The ages of the subjects ranged from 18 to 40. None of the subjects reported any speech or hearing impairment. All subjects took part on a voluntary basis and received no remuneration for their service.

4.1.1.3. Procedure

Subjects participated in the experiment in individual sessions. During the experiment the listener was seated in a quiet (but not sound-treated) room, with no other persons present. Subjects were instructed, interactively and in writing, to monitor the non-words that would be presented to them, and to decide for each non-word, as fast as they comfortably could while avoiding errors, whether it did or did not contain a pre-specified CVC sequence. If the stimulus contained the target sequence, the subject was to press a green key on the keyboard of a notebook computer with the right hand; if the non-word did not contain the target sequence, the subject was to press a red key with the left hand.[6]

First the subject heard a target (CVC sequence), followed by a 500ms pause, which in turn was followed by a non-word. A 2000ms time-out interval followed the stimulus. If the subject responded before 2000ms had elapsed, the presentation of the next target was initiated 100ms after the issuing of the response. If the no key was pressed, the next target was presented immediately after the 2000ms time out.

The 192 target-non-word combinations were presented without any break, preceded by 8 practice sequences. A random order of presentation was used for each listener. Stimuli were converted from digital to analog format on a notebook computer and presented to the listeners over high-quality headphones. Stimulus presentation and response collection were fully automated, using E-prime software.[7]

[6] In the Georgian educational system right-handedness is normative. Assuming that no natural differences between right-handed and left-handed subjects exist in this population, the subjects were instructed to use their preferred hand for positive responses.

[7] The script for the experiment was written by Ing. Jos J.A. Pacilly of the Leiden University Phonetics Laboratory.

4.1.2. Results

In total 24 (subjects) x 192 (stimuli) = 4,608 responses were collected. Only responses to target-bearing stimuli were statistically analysed, reducing the nominal number of responses to 2,304. Hit rate and mean latencies (for hits only) were computed and subjected to one-way analyses of variance (ANOVA) with consonant type (harmonic Cc, unrelated CC, single C) as a fixed effect. Post hoc t-tests for contrasts between pairs of consonant types were run if the overall effect was significant.

In the following subsections, the results for the dependent variable 'hit rate' are presented, and then the 'latency' data are considered. The data are also analysed, after selection of a specific, phonologically (distributionally) motivated, subset with coronal stops and affricates only.

4.1.2.1. Hit rate

The table in (7) presents means and standard deviations for the dependent variable 'hit rate', broken down by the three types of consonant sequences.

(7) *Mean hit rate (%), standard deviation and number of observations, broken down by type of stimulus sequence*

Type of stimulus sequence	Mean	SD	N
Unrelated (CC)	86	35	754
Single (C)	78	41	748
Harmonic cluster (Cc)	80	40	748

The table in (7) shows that the CVC target was correctly detected in stimuli where the target was an integral syllable (unrelated CC, i.e. the easiest condition) in 86% of the cases. The hit rate is substantially poorer (80%) in the harmonic cluster (Cc) context, which is as predicted. However, the hit rate in the single consonant (C) context, which we predicted to assume an intermediate position along the difficulty scale, was poorer still (78%). The overall effect is significant, $F(2,2247) = 6.8$ ($p = 0.001$). The post hoc tests reveal that unrelated CCs differ from the other two types, which do not differ from one another.

Several studies have claimed a difference within harmonic clusters on distributional grounds: coronal + dorsal vs. labial + dorsal (Vogt 1958, Chikobava 1971 and Deprez 1988). Firstly, coronal + dorsal harmonic clusters may be part of three-member clusters, while labial + dorsal harmonic clusters may not. Secondly, coronal + dorsal harmonic clusters participate in long consonant sequences, while the labial + dorsal type does not. On the basis of these distributional criteria, Deprez (1988) proposes that all coronal + dorsal harmonic clusters should be treated as complex segments, whilst the labial + dorsal type should be analysed as either true (CC) clusters or as complex (Cc) segments, depending on some lexical distinction.

Since all the stimuli are non-words, there is no way of knowing which labial + dorsal Cc clusters should be analysed as true CC clusters. Therefore the analysis was

repeated after selecting only responses to stimuli that contained the coronal + dorsal type (for which the predictions are clear cut). To be on the safe side, the harmonic types that began with a continuant (these seem to pattern with the labial + dorsal type) were also excluded. As a result of this data selection, only the types contained in rows 2, 3 and 4 in (1) were retained in the analysis. This left us with 18 harmonic clusters (instead of 32) and 54 stimuli. The table in (8) presents the breakdown of results.

(8) *Mean hit rate (%), standard deviation and number of observations, broken down by type of stimulus sequence. Only cluster types with coronal stops and affricates were selected.*

Type of stimulus sequence	Mean	SD	N
Unrelated (CC)	86	35	446
Single (C)	74	44	442
Harmonic cluster (Cc)	75	34	443

The results are qualitatively the same as in the table in (7), but the scores for the harmonic clusters (Cc) and those of the single consonants have dropped a few points, so that they differ more from the unrelated C + C sequences. Consequently, the overall effect is more strongly significant, $F(2,1328) = 10.7$ ($p < 0.001$). CC differs from both Cc and C, which do not differ significantly from each other.

To conclude, separating the coronal obstruent cases does not substantially change the results. Therefore, there seems to be no firm experimental support for the claim that complex segment status applies only to coronal + obstruent clusters.

4.1.2.2. Reaction time

The table in (9) presents mean detection time (in milliseconds) for hits only, broken down according to the type of cluster. Detection time was defined as the time interval between the onset of the stimulus sequence and the moment the listener pressed the response key.

(9) *Mean and standard deviation of target detection time (ms) for hits only, measured from the stimulus onset, and number of observations, broken down by type of stimulus sequence*

Type of stimulus sequence	Mean	SD	N
Unrelated (CC)	1041	354	645
Single (C)	1007	335	587
Harmonic cluster (Cc)	1052	362	600

The results indicate that the target was detected somewhat faster for the single C condition than for the other two conditions, which do not seem to differ from each

other significantly. However, the effect is a trend at best, as the ANOVA fails to reach significance; $F(2,1829) = 2.7$, $p = 0.067$ (ins.).

Of course, the listener can only detect the target once the third segment in the stimulus sequence has become available. Since there may well be a difference in temporal structure between the three types of stimuli, the detection time had been redefined as starting from the earliest moment in the acoustic signal of the stimulus that fully contained the CVC sequence specified by the target. That is to say, the reaction time was measured from the offset of the first (or only) consonant in the crucial sequence. The results for the corrected detection times are presented in the table in (10).

(10) *Mean and standard deviation of target detection time (ms) for hits only, measured from the offset of C1 in stimulus, and number of observations, broken down by type of stimulus sequence*

Type of stimulus sequence	Mean	SD	N
Unrelated (CC)	894	389	645
Single (C)	844	372	587
Harmonic cluster (Cc)	892	407	600

The results are qualitatively the same as before, but the difference between the relatively fast detection time for the single C condition and the remaining two has increased to about 50ms, and reaches significance, $F(2,1829) = 3.2$, $p = 0.042$. Single Cs differ significantly from the other conditions, which do not differ from each other (Scheffé test).

On the basis of the table in (10) then, we conclude that the harmonic clusters pattern with the unrelated CC sequences. This patterning is different from that found in the results for hit rate (the tables in (7) and (8)).

4.1.2.3. Summary of results

The table in (11) summarises the results of the experiment in terms of relative degree of ease with which the CVC target was detected in the stimuli, broken down according to the three types of stimulus sequence. In the table '+' stands for 'relatively easy detection', whilst '–' represents 'relatively difficult detection'. Of course, easy detection corresponds with a high percentage of correct detections and low detection times.

(11) *Relative ease '+' versus difficulty '–' for hit rate and detection time, broken down by stimulus type*

Stimulus sequence Dependent	Unrelated CC	Harmonic Cc	Single C
Hit rate (%)	+	–	–
Detection time (ms)	–	–	+

The summary shows quite clearly that the harmonic Cc clusters behave as single consonants in terms of hit rate, but as a sequence of two unrelated consonants in terms of detection time. These patters can be interpreted as an indication that harmonic clusters are neither true consonant sequences nor simplex consonants. They have a separate, intermediate, status in the phonology of Georgian. Also, the table in (11) indicates that detection of a CVC sequence was always difficult in a harmonic cluster context. These patterns seem to ascertain that harmonic clusters are indeed complex segments, perceived as a unit by Georgian native listeners.

4.2. Conclusion and discussion

In the introduction to this chapter some straightforward predictions were proposed regarding the detectability of a word-initial CVC target that was embedded in three types of stimulus contexts, viz. the single C condition (CV.CV…), the harmonic cluster condition (CV.CcV…) and the unrelated sequence condition (CVC.CV…). It was predicted that ease of detection would show up simultaneously in high detection rates (hit rates) and fast detection times (for hits). Moreover, it was predicted that the target CVC would be most difficult to detect if the final consonant was the first part of a harmonic cluster in the onset of the second syllable in the stimulus string. The rationale behind this prediction was that the listener would have to break up a highly cohesive phonological unit, i.e. a complex segment, in order to match the target CVC with the stimulus string, and make the match in spite of a difference in syllable boundary.

Although high hit rates usually match fast detection time, this result did not obtain throughout the experiment. Yet, in the crucial condition, i.e. the complex segment condition, the predicted correlation between poor hit rate and long detection time was found. Moreover, the harmonic clusters were consistently found to be the most difficult type in the experiment, whether the results were expressed in terms of hit rate or detection time. The reference conditions (CC and C) were easier to detect than the harmonic clusters in terms of either hit rate (unrelated CC) or detection time (single C). On the strength of this evidence, it can be argued that the perception experiment has substantiated the phonological status of Georgian harmonic clusters as complex segments. Apparently, a perceptual technique, such as target detection in nonsense strings, may provide a useful diagnostic in the determination of the phonological status of some sound sequences when the study of speech production or acoustics yields indeterminate results.

However, before any definitive conclusions, whether substantive or methodological, are drawn from the experiment, the experimental technique should be calibrated against a precise set of contrasts between simplex and complex segments. After all, one might argue that if the Georgian complex segments were truly single consonants, it is hard to understand how the experimental subjects were at all able to parse them into plosive + velar obstruents. Although the task of matching the target string to a complex segment was more difficult, the subjects were still quite successful in accomplishing their task. It would make sense, therefore, to run a similar experiment in English, where the affricates /ǰ č/ constitute clear and undisputed examples of complex segments that should be analysed as single units – relying on synchronic distributional criteria, phonological processes and their behaviour in speech errors (affricates are always interchanged with single consonants; Fromkin 1971), as well as their historical origin (Steriade 1989 and van de Weijer 1996, among others). As in Georgian, the 'components' of English complex segments also occur as single consonants in the phonology of the language, so that the task of splitting the complex segment and matching its first component to a single stop should be feasible. The expectation is that the English results will be highly comparable to Georgian data.

5 Reduplicative Patterns in Georgian

5.0. Introduction

One interesting aspect of reduplication phenomenon, the process implying the repetition of all or part of the radical element (base), is that it can be regarded as iconic in nature, or alternatively, as characterised by "self-evident symbolism" (Sapir 1921:79). According to Jakobson (1949), reduplication is an iconic relation between the linear morpheme order and the meaning which the word denotes (e.g. plurality, iteration, continuance, increase in size, repetition and duration).[1] The peculiar type of form-meaning mapping in reduplication requires a specific phonotactic organisation of the reduplicated form. Iconicity yields predictability, which is manifested in transparent phonological structure. Transparency is related to the notion of simplicity, which is a characteristic feature of unmarked structures in a language.

Reduplication is a type of affixation process which takes into account the phonology of the base (the reduplicant). Cross-linguistic study of reduplication has shown that reduplicants are less marked in particular ways than the remainder of the outputs of a language, e.g. in Optimality Theory this is captured by the 'emergence of the unmarked' ranking (McCarthy & Prince 1994, 1995, Carlson 1997, among others). Since reduplication generally yields an unmarked subset of what a language may allow in the way of prosodic and featural structures, reduplicative patterns of Georgian are discussed to substantiate several claims made in Chapter 3: (i) the minimal word in Georgian is disyllabic; (ii) harmonic clusters can be analysed as complex segments; (iii) C + /v/ combinations can be analysed as complex segments; (iv) /s/ + obstruent can be analysed as complex segment; (v) obstruent + sonorant is the most unmarked consonant cluster; and (vi) Georgian maximally allows two-member consonant clusters stem-initially. Consequently the data support the claim that consonant sequences are, in general, derived.

There are numerous compounds in Georgian that are formed by reduplication. Reduplication occurs in verbal forms, e.g. to mark continuative aspect or intensity and in nominal forms, e.g. to form the collective plural. Reduplication is widely used in onomatopoeia, too. The reduplicant commonly attaches to a base as a suffix. Total reduplication, which involves copying of a complete base, is quite common. Partial reduplication is also found. It obeys the general co-occurrence restrictions of

[1] Note that not all types of reduplication are strictly iconic, e.g. in Leti reduplication is used to form the agentive of nouns, e.g. *sopan* 'message', but *sopsopan* 'messenger'.

the language. All the reduplication data discussed in this chapter come from the extensive study of verbal and nominal reduplication by Ertelishvili (1970, 1980).

In the following sections, reduplication data are presented in order to reveal the patterns of the minimal word size in Georgian and of segmental patterns.

5.1. Prosodic patterns

It has been pointed out that reduplication patterns seem to indicate that certain prosodic structures of a language are unmarked. In Chapter 3, I proposed a disyllabic minimality constraint on Georgian prosodic words. The syllabic constituency of the reduplicated forms is interesting with respect to this issue.

The base of reduplication can be either a *prosodic word* or a *stem* in Georgian. Word reduplication differs from stem reduplication in that semantically and prosodically different types of derived forms are created as a result of the copying procedure. The differences are illustrated in the following sub-sections.

5.1.1. Word reduplication

There are no monosyllabic words that participate in reduplication processes, e.g. forms like **da-da* 'sister', **tve-tve* 'month', **xe-xe* 'tree' are ill-formed.

Only disyllabic words can reduplicate in Georgian. Consider the case of the reduplication of nominal forms expressing the collective plural.

(1) *Complete word reduplication*

kuča-kuča	'street to street'
vak'e-vak'e	'lowland, valley'
k'ona- k'ona	'bunch of flowers'

Trisyllabic words are ill-formed bases for reduplication.

(2)

*maɣali-maɣal-i	>	maɣal-maɣal-i	'tall'
*lamazi-lamaz-i	>	lamaz-lamaz-i	'beautiful'
*axali-axal-i	>	axal-axal-i	'new'

The word-final vowel /-i/, which is the nominative case marker, does not appear in the base, but is retained in the reduplicant since, according to the phonotactic constraints of Georgian, no word can end in a consonant. Thus, trisyllabic words do not form bases for reduplication.

To conclude, neither monosyllabic nor trisyllabic words are appropriate bases for reduplication. Only disyllabic words are suitable bases, as presented in the examples in (1). Furthermore, note that in the case of total reduplication compounds are formed with a primary accent both on base and reduplicant as shown in (3).

(3) *The accentual pattern of word reduplication*

kúča-kúča	‘street’	k’óna-k’óna	‘bunch of flowers’
vák’e-vák’e	‘plain’	lámaz-lámaz-i	‘beautiful’

As mentioned earlier, word reduplication differs from stem reduplication. The difference can be seen from accentual patterns and I will return to this issue in the following section.

An interesting argument for the disyllabic minimality constraint on Georgian prosodic words comes from reduplicated forms with a vowel alternation. There are many cases of *a* (base) ~ *u* (reduplicant) alternation with the epenthetic vowels /a/ or /i/ after stems of the C(C)VC type.

(4) *Cases with the epenthetic vowel* /a/

bak’a-buk’-i	‘to make a noise with heels’
lac’a-luc’-i	‘crash, crack’
čxara-čxur-i	‘tickling’
txlaša-txluš-i	‘to slap, smack’
baχ’a-buχ’-i	‘buttering, mumbling’
k’ak’a-k’uk’-i	‘to knock’
pacxa-pucx-i	‘fuss, bustle’

(5) *Cases with the epenthetic vowel* /i/

panci-punc-i	‘quivering’
parti-purt-i	‘fuss’
lac’i-luc’-i	‘crack’

The examples illustrated in (4) are (5) are phonologically and prosodically much like those in (1). The insertion of the epenthetic vowels /a/ or /i/, i.e. the extension of stems (bases) with final vowels, should be regarded as satisfaction of the well-formedness condition on Georgian prosodic words, which are always disyllabic. The choice of the vowel seems random, although cases with the epenthetic vowel /a/ are the most common.

5.1.2. Stem reduplication

The Georgian stem is commonly of the C(C)VC type and reduplication takes this sequence as the base. Such reduplicated forms have the nominative case suffix /-i/ to satisfy the word-final requirement of having an open syllable. The most commonly attested pattern of complete reduplication is that of stems of the CVC type.

(6) *Complete CVC stem reduplication*

baχ'baχ'-i	'to speak in a deep voice'
buχ'buχ'-i	'mumbling'
taxtax-i	'regular, slow movement'
ǰagǰag-i	'to go at a jog-trot'
puspus-i	'fidgety'
c'ak'c'ak'-i	'tickling'
čakčak-i	'to go at a jog-trot'

The reduplicated forms illustrated in (3), (4) and (5) differ from the examples in (6) in that the forms in (3), (4) and (5) are compounds containing two words, while the forms in (6) are single derived words. The difference is expressed orthographically as well. The forms in (3), (4) and (5) are hyphenated, while the forms in (6) are not. Furthermore, the two types of word-forms differ in their accentual patterns: the words in (6) form one prosodic word with a primary accent on the first syllable, e.g. *táxtaxi*, *púspusi*, while the words in (3), (4) and (5) contain two prosodic words with two primary accents on the initial syllables, e.g. *kúča-kúča*, *lámaz-lámaz-i*, *čxára-čxúr-i*, *pánci-púnc-i*.

Other types of stem reduplication are discussed in the following sections when consonantal patterns are considered.

5.1.3. Conclusions

In this section it has been shown that Georgian prosodic words can reduplicate totally when they meet the condition of disyllabicity. In non-disyllabic words, vowel deletion or epenthesis is applied in order to meet the disyllabic constraint (see (2), (4) and (5)). To conclude, the reduplicative patterns discussed in this section support the claim made in Chapter 3 that the Georgian minimal prosodic word is disyllabic. Note that the claim does not exclude the existence of monosyllabic words in Georgian, e.g. /xe/ 'tree', /da/ 'sister', /dɣe/ 'day', but treats them as marginal to the language system.

5.2. Segmental patterns

Since this thesis focuses mainly on consonant sequences, only consonantal patterns of reduplicated forms are considered in the following sections. The reduplicated forms are interesting with respect to the status of harmonic clusters and C + /v/ combinations. In Chapter 3, I argued that Georgian maximally allows two-member consonant sequences, predominantly of the obstruent + sonorant type. They usually appear stem-initially. Georgian avoids consonant sequences in stem-final position. Long consonant sequences are derived, marked structures of the language. All these claims are examined in the light of reduplication data.

5.2.1. Harmonic clusters

Ertelishvili's (1970, 1980) data on the reduplication of verbal roots in Georgian contain numerous examples of harmonic clusters participating in the reduplication process. Most of the examples are onomatopoeic. The reduplicated CV sequences, and sometimes CVCV sequences as well, have parallel forms in which a velar or post-velar obstruent is inserted after the second consonant. Thus, reduplicated forms of the CVCCV type are attested. The reduplication of the CV sequence is illustrated in (7).

(7) *The reduplication of the CV sequence and insertion of dorsals*

sisin-i[2]	sisxi-n-i	'hissing'
t'at'an-i	t'at'χ'a-n-i	'speaking loudly'
cicin-i	cicki-n-i	'yelping'
č'ič'in-i	č'ič'χ'in-i	'speaking piercingly'

As a result of reduplication, the harmonic clusters *sx*, *t'χ'*, *ck* and *č'χ'* are created in the reduplicant. The phenomenon could also be viewed not as insertion of dorsals, but as alternation of consonants *s* ~ *sx*, *c* ~ *cx* and *č'* ~ *č'χ'* as a result of which the parallel forms, given in (7) emerge, e.g. *sisin-i* ~ *sisxi-n-i*, etc. Harmonic clusters substitute single consonants without any change in meaning. These examples are good evidence for the claim that harmonic clusters behave as single segments. Note that bases for reduplication in these examples are monosyllabic, e.g. *si* in *sisini* or *t'a* in *t'at'an-i*, but they are not independent words and thus cannot challenge the disyllabic constraint advocated in 5.1.1.

Parallel forms with a single consonant/harmonic cluster are attested in onomatopoeic words formed by reduplicating the base of the CVC type, e.g.

(8) *CVC reduplication with harmonic cluster substitution*

zurzul-i[3]	zurzγul-i
tantar-i[4]	tantkar-i
c'inc'il-i[5]	c'inc'k'il-i

There are interesting cases of parallel forms where harmonic clusters appear not as a result of substitution of a single consonant, but as a result of vowel deletion in the reduplicant. The process accompanies the change from a closed (CVC) syllable type to an open (CV) syllable.

---

[2] The sonorant /n/ often appears in infinitival forms. Its behaviour and status is marginal to the discussion here and I leave it unaccounted for.

[3] Often the sonorant /r/ of the base is substituted by the sonorant /l/ in the reduplicant, e.g. reduplicated forms such as *bar-bal-i*, *t'ar-t'al-i*, *kir-kil-i*, *gur-gul-i*, etc. are well attested.

[4] The reduplication of the sonorant /n/ of the base is disallowed in the reduplicated part. It is replaced either by the sonorant /r/ or /l/, e.g. *sa**n**-sa**l**-i*, *χ'a**n**-χ'a**l**-i*, *su**n**-su**l**-i*, *k'u**n**-k'u**l**-i*, *ta**n**-ta**r**-i*, etc.

[5] The onomatopoeic words formed by means of reduplication are numerous. It is difficult to find an adequate translation for such words. All examples that are not glossed in this chapter are onomatopoeic.

(9) *Vowel deletion and harmonic cluster formation*

a)	jagjag-i	>	b)	jajga-n-i	‘speaking angrily’
	sixsix-i	>		sisxi-n-i	‘hissing’
	jiɣjiɣ-i	>		jijɣi-n-i	‘speaking angrily’

The forms in (9) illustrate the change to an unmarked structure in two aspects:

(10) a) In terms of the syllable type: from closed to open (which is more widely attested cross-linguistically), e.g. *ja.jga.ni* and *si.sxi.ni*.
b) In terms of the segmental content of the word: word-medial consonant sequences (e.g. *gj*, *xs* and *ɣj* in (9a)) are simplified and changed into harmonic clusters (*jg*, *sx* and *jɣ* in (9b)).

Besides the emergence of harmonic clusters in reduplicative forms, it is remarkable that harmonic clusters are the only obstruent sequences to reduplicate. Thus, in the base of the OOVC (O = obstruent) type only a harmonic cluster can replace OO.[6]

(11) *The reduplication of stem-initial harmonic clusters*

tkar-tkar-i	c’k’av-c’k’av-i
tkon-tkor-i	sk’up’-sk’up’-i
txap’-txap’-i	dgan-dgar-i
t’χ’up’-t’χ’up’-i	

It was shown in Chapter 3 that harmonic clusters are special in terms of their distributional patterns, since they are the only obstruent clusters which can appear in stem-final position. There are reduplicated forms illustrating harmonic clusters with this pattern.

(12) *The reduplication of stem-final harmonic clusters*
račxa-ručx-i
pacxa-pucx-i
razɣa-ruzɣ-i

The remarkable behaviour of harmonic clusters in reduplication processes supports their analysis as complex segments. If one accepts the equation of markedness of a structure with its complexity, then the data given in (7) to (12) could be used as an argument for the analysis of harmonic clusters as complex segments, as advocated in Chapter 3.

[6] There is no case of reduplication of the base of the CVOO type.

5.2.2. C + /v/ sequences

In the previous chapter, I argued that C + /v/ combinations are complex segments. To support this claim, I consider reduplication patterns similar to those demonstrated for harmonic clusters in section 5.2.1.

There are many cases of labialisation of the first consonant of the reduplicant. These consonants may be dental, alveolar or velar, but never bilabial. This is in accordance with the general phonotactic patterns of Georgian, which exclude combinations of two bilabials.

(13) *Labialisation in the reduplicant*

sirsval-i	cercvel-i
čirčval-i	ɣirɣval-i

Reduplication of the base of the C/v/VC type is also quite common.

(14) *The reduplication of C + /v/ combinations*

k'vink'vil-i	c'k'varc'k'val-i
ɣvanɣval-i	xvarxvar-i
xvanxval-i	χ'vanχ'val-i
kvankval-i	gvangval-i
ɣvarɣvar-i	čvarčval-i

Notice that most of the examples of this type of reduplication involve velar obstruents as landing sites for labialisation. The same effect is operative in the metathesis process (see Chapter 3).

These two reduplication patterns (see (13) and (14)) suffice to demonstrate that in an environment when the sonorant /v/ is preceded by a consonant (especially by a velar), it serves as a secondary articulation on the consonant. The C + /v/ combinations, like harmonic clusters, are treated as complex segments by the reduplication process. Note that other types of consonant sequences do not participate in reduplication process (see section 5.2.5).

To conclude, according to the reduplicative patterns of the C + /v/ combinations, the sequences can be treated as complex segments, as proposed in Chapter 3.

5.2.3. /s/ + obstruent sequences

Some of the clusters of the /s/ + obstruent type, e.g. *sx* or *sk*, have already been discussed. These are harmonic clusters, and are treated as complex segments. There are also a number of /s/ + obstruent clusters which are not harmonic clusters, but which nevertheless behave like complex segments and could be analysed as such. These clusters, unlike other types of clusters (e.g. *rk'*, *rj*, *ǰd*, *k'b*, *gd*, etc.), participate in the reduplication process.

(15) sc'orsc'ori 'straight'
sk'up'sk'up'i 'to jump'
st'ven-a-st'ven-it 'whistling'

The similar behaviour of /s/ + obstruent sequences in reduplicative contexts is observed in a number of languages, e.g. in Gothic and Sanskrit. The special behaviour of /s/ + obstruent sequences led to the proposal that they can be treated as complex segments (e.g. Selkirk 1982, Broselow 1991, van de Weijer 1996).

To conclude, there are three types of consonant sequences, harmonic clusters, C + /v/ combinations and /s/ + obstruent clusters, which participate in the reduplication process and behave as complex segments.

5.2.4. Sonorants

The most common type of reduplication is that of CVC stems which have a sonorant as the second consonant (Ertelishvili 1970).

(16) *Sonorants as the second consonant of the base*
gurgur-i 'buzz'
tamtam-i 'slow movement'
zorzor-i 'noise'
p'arp'al-i 'blistering'
k'amk'am-i 'shine, beam'
čurčul-i 'whisper'

Examples of sonorants as the first consonant of the base are as follows:

(17) *Sonorants as the first consonant of the base*
rat'rat'-i laklak-i
recrec-i višviš-i

In most of these cases the sonorant is either /r/ or /l/.

The type of reduplication in (16), unlike that in (17), can entail vowel reduction in either the reduplicant or the base.

(18) *Vowel deletion and consonant-cluster formation*
čirčil-i > črčil-i 'moth'
gorgol-i > grgol-i > rgoli 'ring'
parpal-i > parpl-i 'fin'

Note that reduction results in consonant clusters of rising sonority, *čr*, *gr*, *pl* and *kl*, which, as I argue, are the least marked consonant-sequence types in Georgian.

There are cases of complete reduplication involving the sequences with consonant + sonorant as their initial elements.

(19) *The reduplication of consonant + sonorant sequences*
tlaxtlax-i
c'rup'c'rup'-i
kniškniš-i

The reduplicative patterns of the consonant + sonorant type illustrate that they are well-formed consonant sequences in Georgian.

5.2.5. Consonant sequences

The consonantal pattern of reduplication shows that Georgian can reduplicate bases which have, at most, two-member sequences. These sequences are harmonic clusters, which are sometimes extended by the bilabial sonorant /v/ (as in the case of *c'k'va-c'k'val-i*) or by the clusters of the consonant + sonorant type. In the former case, if the harmonic and the C + /v/ clusters are treated as complex segments, as suggested in Chapter 3, then the analysis of the surface three-member cluster *c'k'v* is that it a complex segment, i.e. the labialised segment *c'k'*^{w}. There are also /s/ + obstruent clusters which are treated as complex segments. All of these considerations leave us with consonant + sonorant sequences as the only true two-member clusters in Georgian.

The reduplication data discussed in this chapter support the claims made in Chapter 3 that two-member clusters are the longest true clusters in Georgian. In addition to this observation, the reduplication data illustrate that there are no cases of reduplication in which a long consonant sequence participates. Note that long consonant sequences occur in words, e.g. /brj/ in /brjen-i/ 'wise' and /mk'vr/ in /mk'vriv-i/ 'dense', but do not participate in any reduplication process. Examples of consonant-sequence simplification would substantiate the claim (**brjen-jen-i* or **mk'vriv-k'vriv-i*), but to my knowledge there are no examples of this type. Reduplication simply does not apply to words or stems with long consonant sequences. Perhaps this is an indication that stems or words that have synchronically long sequences are, in fact, morphologically complex. As we have seen in this chapter, reduplication takes only a monomorphemic base. The failure to reduplicate these long consonant sequences, which do not obey the SSP and are longer than two consonants, could be used as an argument for their marked, derived character. This is in accordance with the claim made in Chapter 3 that these consonant sequences are derived.

In addition to constituency patterns, reduplication reveals the markedness of the distributional patterns of consonant clusters. As has already been mentioned, clusters are commonly found in base-initial position, while base-final clusters other than harmonic clusters are excluded. The distributional patterns of consonant clusters also show that the longest clusters are word/stem-initial in Georgian.

5.3. Conclusions

Reduplication data have been considered to reveal the unmarked prosodic and segmental structure of Georgian. Reduplicative patterns support the claims made in Chapter 3: (i) the minimal word is disyllabic; (ii) harmonic clusters can be analysed as complex segments; (iii) C + /v/ combinations can be analysed as complex segments; (iv) /s/ + obstruent can be analysed as complex segment; (v) obstruent + sonorant is the most unmarked consonant cluster; and (vi) Georgian maximally allows two-member consonant clusters stem-initially. Consequently, the data support the claim that consonant sequences are, in general, derived.

6 The Gradual Consonant Analysis

6.0. Introduction

Many linguists have noticed that long consonant sequences occur predominantly in verbal rather than in nominal forms in Georgian (Vogt 1961 and Ertelishvili 1970, among others). This could be related to the fact that Georgian verbs, as has already been mentioned in Chapter 3, have a much more complex morphological structure than nouns. Thus, verbal forms take up a larger part of the discussion in this chapter, although analyses of nominal forms are also considered. This chapter contains two sections. In section 6.1, consonantal verbal stems are examined. Comparative and historical evidence suggests that Modern Georgian consonantal stems containing up to four consonants are historically derived from stems that have a vowel between the consonants. Section 6.2 introduces the Gradual Consonant Analysis. Its application is demonstrated on verbal and nominal forms containing long consonant sequences.

6.1. Consonantal stems

The data considered in this section are meant to prepare the reader for the analysis, the Gradual Consonant Analysis, which follows. A close study of the verbal forms suggests that consonant sequences in Modern Georgian are historically derived from much simpler forms. They are the result of processes of vowel deletion and complex segment formation.

Ertelishvili (1970), in his study of the phonotactic structure of verbal stems, suggests that consonant sequences, their history and characteristics are directly related to the phonotactic patterns of stems. He illustrates this by examining consonantal verbal stems. Modern Georgian stems that consist only of consonants are derived from Old Georgian stems that have a vowel between the consonants. Sometimes the corresponding stems with a vowel between the consonants are also attested in paradigmatically related forms and in dialectal forms.

Consider the following verbal stems, consisting of two, three and four consonants. The data are taken from Ertelishvili (1970). According to Ertelishvili (1970), there are 45 biconsonantal verbal stems in Georgian. The number of consonants is counted as they appear in surface forms. I consider some of the types of biconsonantal verbal stems and illustrate that in most cases consonant sequences are derived after stem-vowel deletion has applied.

(1) *Stems with two consonants*

	Modern Georgian		*Old Georgian*	
Stems	CC		CVC	
bn/ben	e-u-bn-eb-a	‘somebody says to somebody’	v-e-u-ben-i-t	‘we spoke’
sm/sem	e-sm-i-s	‘somebody hears’	še-g-e-sem-in	‘you will hear’

In the examples in (1), the stem vowel deletes when the stem is preceded or followed by a grammatical affix containing a vowel.

(2) $(V)_{affix}$-CVC_{stem}-$(V)_{affix}$ > $(V)_{affix}$-CC_{stem}-$(V)_{affix}$

When the second consonant of the stem of the CC type is /v/ in Modern Georgian, the corresponding form in Old Georgian has the vowel /u/ instead of /v/.

(3) *Stems of the C/v/ type*

	Modern Georgian	*Old Georgian*	
Stems	C/v/	CV	
sv/su	še-sv-a	da-su-a	‘somebody made sit’
c’v/c’u	da-c’v-a	da-c’u-a	‘somebody burned’

Notably, it seems to be the case that when a stem of the CC type contains a harmonic cluster, relatively little diachronic change takes place and harmonic clusters remain unaffected. The clusters behave as complex segments. Consider the following correspondences:

(4) *Stems of the Cc type*

	Modern Georgian		*Old Georgian*	
Stems	Cc		Cc	
cx	a-cx-ob-s	‘somebody bakes’	še-a-cx-o-t	‘you to bake’
c’χ’	da-a-c’χ’-o	‘somebody had put’	še-u-c’χ’-ev	‘put you’

There are also biconsonantal stems in Modern Georgian which do not correspond to Old Georgian stems of the CVC type; however, comparison of such consonantal stems with corresponding forms in other Kartvelian languages, e.g. Svan and Megrelian, provides some evidence for their derived nature. One of these stem types is the sequence C/r/, which has a corresponding stem of the CV/r/ type in other Kartvelian languages.

(5) *Stems with /r/*

	Modern Georgian	*Svan*	*Megrelian*	
Stems	Cr	CVr	CVr	
br/bVr	br-ial-i	bər-bər	bar-bač	‘sparkle’
gr/gVr	gr-ial-i	gər-gən	gur-gun	‘rumble’
zr/zVr	zr-ial-i	zər-zən	zir-zol	‘humble’
pr/pVr	pr-ial-i	pər-pən	par-pal	‘to fly’

As shown in the examples in (5), Modern Georgian stems of the C/r/ type correspond to stems in other Kartvelian languages which have a vowel between the consonants, i.e. CV/r/. In Svan, the stem vowel is a schwa, and in Megrelian, the vowel can be /a/, /u/ or /i/.

There are 42 stems in Georgian containing three consonants only. As illustrated below, these consonant sequences emerge as the result of stem-vowel deletion. The process, as already suggested in (2), occurs in the following context: $(V)_{affix}$-CVC_{stem}-$(V)_{affix}$.

(6) *Stems with three consonants*

	Modern Georgian		*Old Georgian*	
Stems	CCC		CVCC	
drk’/derk’	drk’-eb- a	‘somebody bows’	mo-v-derk’	‘I bowed’
k’rt/k’ert	k’rt-om-a	‘to tremble’	gan-h-k’ert	‘you trembled’

Once again, as in the case of the biconsonantal stems in (3), Modern Georgian /v/ corresponds to the vowel /u/ in Old Georgian.

(7) *Stems with /v/*

	Modern Georgian		*Old Georgian*	
Stems	C/v/C		CuC	
xvd/xud	še-xvd-a	‘he met somebody’	še-m-xud-a	‘somebody met me’
	CC/v/		Ccu	
tkv/tku	tkv-a	‘somebody said’	tku	‘said’

Similarly to the examples in (4), harmonic clusters behave as complex segments.

(8) *Stems with Cc*

	Modern Georgian		*Old Georgian*	
Stems	CCc		CVCc	
rtx/ratx	gan-i-rtx-a	‘was spread’	gan-i-ratx	‘spread you’
rt’χ’/rat’χ	še-rt’χ’-m-a	‘to bound’	še-i-ratχ’	‘somebody bounded’

According to Ertelishvili (1970), stems with the sonorant /r/ between identical consonants are always derived from reduplicated forms.

(9) *Stems with /r/*

	Literary Georgian		*Dialectal form*	
Stems	C_irC_i		CVrCVC	
trt	trt-ol-a	‘to tremble’	tar-tal-eb-s	‘tremble’
	Literary Georgian		*Related form*	
prp	prp-en-a	‘to adore’	par-pat’-i	‘to fly around’
			pr-en-a	‘to fly’

Note that the forms of the Modern Georgian C/r/ stems given in (5) are also derived from reduplicated forms.

As was argued in Chapter 3, the sonorant /r/ is syllabic in two contexts: (i) when preceded and followed by consonants which do not share laryngeal specification, and (ii) when preceded and followed by identical consonants. The examples in (6) and (9), where the Modern Georgian forms have corresponding forms (in (6) Old Georgian and in (9) Svan and Megrelian) with a vowel before the sonorant /r/, also substantiate the claim that Modern Georgian /r/ is syllabic in those contexts.

Stems containing four consonants only are very rare, and similarly to the previous cases, it is always possible to reconstruct a vowel between the consonants by taking into consideration paradigmatically related forms, historical and comparative evidence. The following alternations are attested:

(10) *Stems with four consonants*

crcv/jarcu	‘to fade’
crcn/crcen/cercen	‘to peel off’
rc’χ’v/c’χ’v/c’χ’ev	‘to deplete’

As shown by the examples in (1), (5), (6) and (9), the comparative, historical and related forms demonstrate that Modern Georgian consonantal verbal stems are derived from stems of the CVC type. The deletion of a stem vowel is triggered by the addition of a vowel-initial affix to the stem. Thus, deletion occurs in the following context:

(11) *The context of stem-vowel deletion*
$(V)_{affix}$-CVC_{stem}-$(V)_{affix}$

The presence of either a prefix or a suffix is necessary for deletion to apply.

As shown by the examples in (1), (5) and (6), the stem vowel which is deleted is usually either /a/ or /e/. Recall from Chapter 3 that these are the vowels which, in general, undergo deletion in Modern Georgian.

To conclude, the data considered in this section suggest that consonant sequences in Modern Georgian are historically derived from much simpler forms. They are the result of processes of vowel deletion and complex segment formation.

6.2. The Gradual Consonant Analysis

This short excursion into the history and constituency of the consonantal verbal stems is intended as an introduction to the types of evidence (e.g. historical, comparative, paradigmatic and syntagmatic) used in the analysis of consonant sequences, called the Gradual Consonant Analysis (GCA), and to the idea that in order to study and analyse the long consonant sequences, various aspects of the language should be taken into account. In the following pages I spell out what 'various' means. The GCA assumes the prosodic hierarchy (segment - stem + (affix) - word) and the view of word phonotactics in three dimensions proposed in Chapter 2.

Before addressing the procedural aspects of the Gradual Consonant Analysis (GCA), I should explain why the analysis is called 'gradual'.

The adjective 'gradual' implies proceeding or happening step by step or by degree. The term 'gradual', as it is used here, refers to the type of procedural mechanism, which applies step by step to the consonant sequences. The procedure continues in the case of the presence of relevant evidence, eventually affecting the degree of complexity of consonant sequences. Each step in the procedure yields a less complex structure, eventually arriving at the simplest possible form. Without proper evidence the procedure does not proceed further. Thus, the existence of proper evidence is crucial to the mechanism of the analysis. What are the types of evidence considered in the GCA?

(12) *Types of evidence in the GCA*
- a) Paradigmatic and syntagmatic
- b) Historical
- c) Phonetic
- d) Comparative

I will briefly summarise what each type of evidence refers to, and then demonstrate the application of the GCA to the Georgian data.

Paradigmatic and syntagmatic evidence encompasses the following:

a) The ability of consonants to participate in minimal pairs, i.e. their distinctive function.
b) The appearance of consonants in paradigmatically related forms.
c) Distributional patterns, i.e. combinatorics of consonants.

These patterns establish the status of consonants in isolation and in sequence. For instance, this type of evidence is used to establish the status of harmonic clusters as complex segments.

Historical evidence, showing the regular changes between the two stages of a language's development, establishes the status of consonants both in isolation and in sequence. For instance, this type of evidence has been used to establish the status of the sonorant /v/ as a secondary articulation on the preceding consonant.

Phonetic evidence is based on articulatory, acoustic or perceptual study. Such studies can reveal the status of consonants in isolation and in sequence, and, eventually, the result of a phonetic experiment can be incorporated in the analysis. For instance, this type of evidence has been used to establish the status of harmonic clusters as complex segments.

Comparative evidence is based on the direct correspondences between the literary form and the corresponding dialectal form, and between the language form and its corresponding form in genetically affiliated languages. For instance, this type of evidence has been used to establish the syllabic nature of the sonorant /r/.

All these types of evidence have been taken into account in this thesis in order to uncover the status of consonants and consonant sequences in Georgian. As a result of the discussions in Chapter 3, 4 and 5, the following generalisations have been proposed and substantiated.

(13)
a) Harmonic clusters are complex segments.
b) Sequences of the C + /v/ type are complex segments.
c) Sequences of /s/ + C are complex segments.
d) The sonorant /r/ is either optional or syllabic in consonant sequences. It is optional when preceded and followed by consonants sharing laryngeal specification. It is obligatory when preceded and followed either by identical consonants, or by consonants not sharing a laryngeal specification.
e) The minimal word is disyllabic, i.e. of the $C_1V_1C_2V_2$ type.
f) The only genuine clusters are two-member ones and even these two-member clusters are the result of stem-vowel deletion.
g) Consonantal complexity is related to morphological complexity.

Taking all these generalisations into account, I now consider the application of the GCA to Georgian consonant sequences.

Consider, for instance the longest possible cluster *prckvn* in *prckvna* 'to peel'. Several generalisations and assumptions should be taken into account. For expository purposes the analysis is presented in a stepwise fashion, the succession of the

steps being fairly random. In the first step (i), the optional appearance of the sonorant /r/ is taken into account. This leaves us with the five-member sequence *pckvn*. In the second step (ii), the harmonic cluster *ck* is treated as one segment. This leaves us with the four-member sequence *p*$\widehat{ck}$*vn*. In the third step (iii), the C + /v/ combinations are regarded as one segment. This yields the three-member sequence *p*$\widehat{ck}^{w}$*n*. In the fourth step (iv), a paradigmatically related form with a vowel before *n* is taken into account, e.g. /ga-v-pcken-i/ ‘I peeled’. Thus, we are left with the two-member cluster *p*$\widehat{ck}^{w}$.

The analysis is depicted in (14).

(14) steps (i)r (ii)$\widehat{ck}$ (iii)C(^{w}) (iv)n → Vn

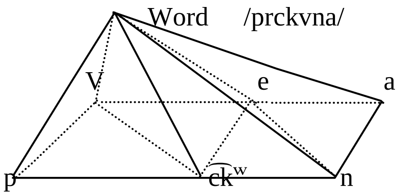
analysis p r c k v n → p c k v n → p$\widehat{ck}$vn → p$\widehat{ck}^{w}$n → p$\widehat{ck}^{w}$

segments CCCCCC → CCCCC → CCCC → CCC → CC

As shown in (14), the analysis of the longest cluster, *prckvn*, yields the two-member cluster *p*$\widehat{ck}^{w}$. One might say that even this two-member cluster can be analysed as a CVC sequence, using comparative evidence. In Megrelian, the corresponding form has a vowel between the consonants *p* and $\widehat{ck}$; thus, the word /purckon-u-a/ ‘to peel’ is attested. This observation leads to the assumption that even the longest Georgian clusters can be analysed as having a CVCV… template.

The word /prckvna/ is a derived word and can be analysed as $[[pV\widehat{ck}^{w}]_{stem}$-$[en]_{affix}[a]_{affix}]_{word.}$ The sonorant /r/, because of its optional occurrence, does not appear in the representation of the word given in (15):

(15)



Thus, the surface six-consonant sequence is the result of vowel deletion, which occurs twice: the first vowel which is deleted is the stem vowel, i.e. the vowel which in the representation in (15) is denoted by the unspecified V, and the second one is a vowel of the affix *–en*, which is deleted after the addition of another infinitival affix, /-a/. Surface complexity is also related to the optional appearance of the sonorant /r/ and complex segment formation (e.g. the harmonic cluster $\widehat{ck}^{w}$). In the metathesis process discussed in Chapter 3 and extensively in Butskhrikidze & van de Weijer (2001a), it is shown that in verbal forms the morphological /v/ tends to become a secondary articulation on preceding consonants, especially the velars. It is possible that in the word *p*$\widehat{ck}^{w}$*na*, the secondary articulation on the harmonic cluster *ck* is derived, i.e. has a morphological origin. It is also possible that it is the result of the

deletion of the vowel /o/ (recall the /o/ ~ /v/ alternation described in Chapter 3). Perhaps the latter suggestion is more plausible, because the corresponding Megrelian form is /purckon-u-a/, with a rounded bilabial vowel /o/ after the harmonic group *c͡kʷ*.

Thus, the word /prckvna/ can be analysed as $[[[[pV\widehat{ck}^{w}]_{stem}\text{-}[en]_{affix}[a]_{affix}]_{word}$, i.e. $[[[[CVC]_{stem}VC_{affix}]V]_{affix}]_{word}$.

Let us now analyse long consonant sequences where the occurrence of the sonorant /r/ is obligatory, for instance the sequence *grgv* in the word *grgvinva* 'thunder'. The application of the analysis to the sequence *grgv* is as follows: in the first step (i), the phonetic syllabicity of the sonorant /r/ is taken into account. In the second step (ii), the C + /v/ combination is regarded as one segment. This yields the sequence *gVrg*^{w}. The procedure is depicted in (16):

(16)	steps		(i) r;		(ii) C^{w}	
	analysis	grgv	→	gVrgv	→	$gVrg^{w}$
	segments	CCCC		CVCCC		$CVCC^{w}$

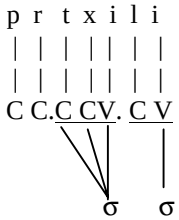
According to Ertelishvili (1970), the form *grgvinva* is historically derived from the reduplicated form **gur-gun*. The form *gur-guli/gur-guri* 'thunder' is attested (the sonorants /n/, /l/ and /r/ alternate quite frequently in reduplicated forms; see Ertelishvili 1970). Thus, the verbal form /grgvinva/ seems to be derived from the noun form by addition of the thematic suffix /-av/ (compare for example the form /grgvin-av-s/ 'thunders') and the infinitival suffix /-a/. The occurrence of the /v/ seems to compensate for the deletion of the bilabial vowel /u/. The /v/ serves as a secondary articulation on the dorsal consonant /g/. Clusters of the *grg* type, such as *trt* in *trtola* 'to tremble' and *crc* in *crcena* 'to fade', are attested in infinitival verbal forms, and as Ertelishvili (1970) suggests, they are derived from the reduplicated forms of the C/v//r/CV sonorant type. These reduplicated forms are usually onomatopoeic words.

Thus, the derived word /grgvinva/ can be represented as $[[[[[gVr]_{stem}[g^{w}in]_{stem}av]_{affix}a]_{affix}]_{word}$.

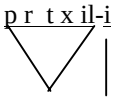
The words /crcena/ and /trtola/ have the same type of representation as the word /grgvinva/.

In Chapter 3, I considered two possible analyses of the word /prtxili/ 'careful'.

(17) a)



b)



Stem + suffix (the nominative case marker)

The schema in (17a) depicts the following patterns of the word /prtxili/: the first two consonants, /p/ and /r/, are not syllabified since the syllable-initial position (onset) is

filled by two consonants (/t/ and /x/). Thus, the word cannot be exhaustively syllabified. We are left with two word-initial consonants (/p/ and /r/), which are not part of the structural constituent called *syllable.* Nevertheless, the whole sequence *prtx* in (17b) is part of a structural constituent *stem* that encompasses the four-member cluster as a whole. Previous analyses on Georgian consonant sequences consider the representation given in (17a) (see Chapter 7, Nepveu's analysis) as more appropriate; however, I consider the representation given in (17b) more appropriate.[1]

Let us analyse the word /prtxili/ 'careful'. Recall that problems arise with regard to the cluster *prtx* and its interpretation in terms of syllable constituency. The cluster *prtx* is a part of the constituent stem in the GCA. The first step in the analysis of the cluster is to take into account the optional appearance of the sonorant /r/. We are then left with the three-term cluster *ptx*. The next step is to consider the *tx* harmonic cluster as a single segment. Thus, we are left with a *p͡tx*, i.e. a CCc cluster. The procedure is depicted in (18).

(18)	steps		(i) r		(ii) Cc	
	analysis	prtx	→	ptx	→	p͡tx
	segments	CCCC		CCC		CCc

The last step would be to consider the comparative method and incorporate dialectal data or direct correspondences from other Kartvelian languages. The form /pərtx/, with a schwa between the /p/ and /t͡x/, is indeed attested in Svan as a direct correspondent to the same root (see Sardschweladse & Fähnrich 1990:321). Thus, the four-term cluster *prtx* is analysed as a sequence of *pVt͡x*, i.e. C_1VC_2. /prtxili/ is a derived word containing the two suffixes *-il* and *-i.* Thus, the representation of the word would be $[[[[[pV\widehat{tx}]_{stem}il]_{affix}]i]_{affix}]_{word}$.

Now let us consider the noun /msxverp'li/ 'sacrifice', with four consonants in word-initial position. In the analysis of this form, several considerations should be taken into account: (i) the sonorant /m/ is a nominaliser in this form, i.e. a prefix, (ii) the sequence /sx/ is a complex segment and (iii) the sonorant /v/ is a secondary articulation on *s͡x*. This leaves us with *s͡x*^{w}. In Old Georgian, the sonorant /v/ had a corresponding vowel /u/. Besides the word-initial consonants this noun form also has word-medial consonant sequences. This word seems to be a compound, i.e. it is morphologically complex.[2]

Consider the adjective /mjlavri/ 'powerful'. As was the case with /msxverp'li/, /m/ is a nominaliser in this form. The consonant sequence *jl* belongs to the stem. Actually the word /mjlavri/ is derived from /jala/ 'power'. Thus, the sequence *jl* is the result of the stem-vowel deletion, which occurs after the affixation of /-av/ to the stem /jala/.

An interesting fact observed in the analysis of forms with long consonant sequences is that words which at the first sight seem morphologically underived are in

[1] For more on the difference in representations depicted in (17) see Chapter 7.

[2] For the time being, I am unable to trace its origin and discuss the association of the consonants /r/, /p'/ and /l/ to any structural constituent.

fact derived. Thus, the formation of consonant sequences and morphological derivation are related processes. This relates directly to the above-mentioned observation that consonant sequences are predominantly found in verbal forms which are characterised by morphologically complex structure.

As has been shown, consonant clustering is primarily the result of stem-vowel deletion, which generally occurs because of the addition of a vowel-initial affix to the stem. Additionally, the optional appearance of the sonorant /r/ and complex cluster formation (e.g. harmonic clusters and labialised consonants) make consonant sequences even more complex. The GCA, by considering different types of evidence, offers in-depth analysis of consonant sequences and predicts where, when and which (in terms of constituency) types of consonant sequences are expected to emerge during language change.

The Gradual Consonant Analysis has confirmed the hypothesis about the derived nature of consonant sequences in Georgian and their relation to the stem domain.

6.3. Consonant sequences and accent

An additional factor directly contributing to the phenomenon of consonant clustering, not considered in this thesis, is accentuation. I have omitted this aspect altogether because much work has still to be done to establish its proper synchronic and diachronic patterns. However, I would like to mention one observation made by Chikobava (1942), which is remarkable with regard to the relationship between the change in accentuation and the process of consonant-sequence formation in Georgian. Chikobava (1942) assumes free accent for earlier stages of Georgian. The placement of accent varied from word to word, i.e. it was lexical. For instance, Chikobava argues that /jma/ 'brother' in Georgian is derived from *jVma*. /jama/ 'brother' is attested in the Western dialect Imeruli; this word also has a direct correspondence in Megrelian as /j;ˇima/. Chikobava (1942) proposes that Modern Georgian consonant sequences are historically derived from structures of the CVCVCV type. As a primary cause of the transformation, he considers the change from free accent (characteristic of Proto-Kartvelian) to fixed accent. Recall that Modern Georgian has fixed accent, which falls on the first syllable of a word (see Chapter 3). The same insight is shared by Machavariani (1957, cited in 1972). He also reconstructs structures of the CVCVCV type in Proto-Kartvelian and, additionally, provides a 'mechanism of accentuation', which is as follows: if accent falls on the first syllable (which is usually the stem vowel), then the affixal vowel is deleted, and if accent is placed on the affixal vowel, then the stem vowel is deleted. The 'mechanism' is discussed in Gamkrelidze & Machavariani (1965). Thus, two structures have been proposed for Proto-Kartvelian.

(19) a) $[CVC]_{stem}$-$[VC]_{affix}$ > $[CVC]_{stem}$-$[C]_{affix}$
b) $[CVC]_{stem}$-$[VC]_{affix}$ > $[CC]_{stem}$-$[VC]_{affix}$

According to Gamkrelidze & Machavariani (1965), the first model (19a) predicts forms such as /der-k'/ 'to bend' and /c'er-d/ 'to write', while the second model (19b) yields forms such as /dr-ek'/ and /c'r-ed/. The former is used in formation of intransitive, past tense verbal forms, and the latter for formation of transitive, present tense verbal forms (Gamkrelidze & Machavariani 1965:370).

The correlations between accent placement, the morphological structure of a word and consonant sequences are interesting, and a synchronic analysis of such correlations is proposed in Butskhrikidze (1998a). Future diachronic study of these correlations should shed more light on the formation of consonant sequences in the Kartvelian languages.

6.4. Conclusions

The examination of consonantal stems of Georgian and the application of the GCA show that in most of the cases long consonant sequences appear to be the result of a very productive vowel-reduction process and complex segment formation. Georgian is an inflectional-agglutinative language, and consonant sequences are generally derived from the CVC stem when vowel-initial affixes are added to it. Thus, phonological 'complexity' is primarily due to morphological complexity. Complex segments also enhance the complexity of consonant sequences.

To conclude, structures of the CCC type in Georgian appear to be the result of transformations of structures of the CVCVCV type, as suggested in Chapter 1.

7 Previous Analyses

7.0. Introduction

In this chapter some recent analyses of Georgian consonant clusters are reviewed, the Headless Syllable Analysis (Nepveu 1994), the Semi-syllable Analysis (Cho & King 1997),[1] the Syllabified Consonant Analysis (Bush 1997) and an analysis within the framework of Government Phonology (Toft 1999). These analyses will be discussed in turn. Comments concerning specific aspects of a particular analysis are made after each section; in 7.4 a comparison of previous analyses with the GCA is made and general conclusions are drawn.

7.1. Nepveu (1994)

Many Georgian consonant clusters seem to violate the Sonority Sequencing Principle and Prosodic Licensing, since they are simply too large to obey both of these constraints at the same time. Nepveu proposes that all such oversized clusters can be broken down into series of sub-clusters of the form T (obstruent), R (resonant) or TR (obstruent + resonant), with at most one sub-cluster occurring at each morpheme boundary. The sub-clusters are the well-formed onsets of headless syllables. Their existence and distribution are accounted for within Optimality Theory and Prosodic Morphology.

Nepveu (1994), as well as Cho & King (1997) and Bush (1997), adopt the following formulation of the SSP and Prosodic Licensing in their analyses of Georgian consonant clusters:

(1) *Sonority Sequencing Principle (SSP)*
Between any member of the syllable and the syllable peak, only sounds of higher sonority rank are permitted (Clements 1990).[2]

---

[1] The Semi-syllable Analysis, proposed by Cho & King (1997), is, however, not discussed here. In Bush's (1997) analysis, the Headless Syllable Analysis and the Semi-syllable Analysis are combined under a single analysis, because the two are quite similar.

[2] The definition of the SSP adopted by the authors of previous analyses is quite dubious. The SSP, as it is formulated in (1), seems to refer only to the onset-nucleus correlation, and does not account for the relation that holds between the nucleus and coda constituent, which is of falling sonority in general. Obviously the SSP is defined with reference to syllable. In Chapter 2 I argued that the domain of the SSP (as well as of the OCP) is not a syllable, but a stem or a word, depending on the language type.

(2) *Prosodic Licensing*
All phonological units must be prosodically licensed, that is, belong to higher prosodic structure (Itô 1989).

Specifically, segments must belong to syllables, syllables to metrical feet, and feet to prosodic words.

Many languages appear to violate the SSP. Clements (1990) therefore argues that the SSP actually holds of underlying syllabification in the lexical phonology and not in surface forms. In addition, Bagemihl (1991), on the basis of data from Bella Coola, challenges the principle of Prosodic Licensing; many segments are proposed to be left unsyllabified in surface forms in this language, e.g. the consonants /q/, /p/ and /s/ in a word such as /qps.ta/ 'to taste'. To account for such forms, Bagemihl (1991) proposes a theory of Moraic Licensing. Nepveu fears that claims made by Clements and Bagemihl threaten the usefulness of the SSP and Prosodic Licensing: "If these constraints may be inoperative throughout a language or operative only at some abstract level of representation, they cannot be used reliably by either the linguist or the language learner in analyzing the surface forms of any given language" (Nepveu 1994:2). He suggests that a close analysis of Georgian reveals that the long sequences of consonants in these languages are far from arbitrary. The sonority contour of these large clusters may rise and fall in apparent disregard of the SSP, but each oversized cluster is composed entirely of sub-clusters that are consistent with the SSP, and there is at most one such sub-cluster at each morpheme boundary. According to Nepveu, this pattern remains unexplained unless there is an organising principle to group the consonants and govern their sonority contours.

Before I present the author's definition of a sub-cluster and its application in the analysis of consonant sequences, some relevant characteristics of the Georgian phonotactics (assumed or proposed throughout his analysis) are given in (3).

(3) a) Only vowels can be syllable peaks in Georgian (Aronson 1990, Robins & Waterson 1952, Vogt 1958). Deprez (1987) claims that /r/ may sometimes be syllabic, but cites only vowels as nuclei.
b) Nepveu adopts Vogt's (1958) analysis of harmonic clusters as complex segments, as well as Deprez's (1987) representation of harmonic clusters with a single root node and a single laryngeal node.
c) Nepveu proposes to analyse the phoneme /v/ as a defective segment, specified only for a labial place of articulation. /v/ acquires other features by fusing with an adjacent segment when possible.

The consequence of this proposal is that /v/ has no inherent values for sonority, voicing or stridency, so that its various phonetic realisations and its ambiguous position in the sonority hierarchy are understandable. "When /v/ fuses with a segment that is already specified as labial, it adds no new information and is invisible. When it fuses with a non-labial segment, it surfaces as an off-glide on the primary articulation" (Nepveu 1994:11).

While addressing the problem of analysing Georgian consonant sequences, Nepveu discusses the following examples:[3]

(4)	a)	zr/d-il-ob-a	/zrd/ ‘raise’, /-il/ PRT, /-ob/ SF, /-a/ NOM
	b)	m/-kr/t-al-i	/m-/ N.AG, /krt/ ‘be afraid’, /-al/ PRT, /-i/ NOM
	c)	r/tm-ev-a	/rtm/ ‘deprive’, /-ev/ SF, /-a/ NOM
	d)	n/gr-ev-a	/ngr/ ‘demolish’, /-ev/ SF, /-a/ NOM
	e)	aɣ-m/-zr/d-el-i	/aɣ-/ PREV ‘up’, /m-/ N.AG, /zrd/ ‘to grow’, /-el/ ADJ, /-i/ NOM
	f)	si-br/jn-e	/si-…-e/ ‘-ness’, /brjn/ ‘wise’
	g)	še-km/n-a	/še-/ PREV, /kmn/ ‘create’, /-a/ NOM
	h)	gr/dem/l-ma	/grdeml/ ‘anvil’, /-ma/ ERG
	i)	c’ig/n-/s	/c’ign/ ‘book’, /-s/ DAT
	j)	sax/l-/s	/saxl/ ‘house’, /-s/ DAT
	k)	semes/t’r-/s	/semest’r/ ‘semester’, /-s/ DAT
	l)	m/-jlav/r-/s	/m-/ N.AG, /jlavr/ ‘powerful, mighty’, /-s/ DAT

While looking at the distributional characteristics of the sub-clusters, Nepveu notes that the general pattern that emerges is the following:

(5) “… the sub-clusters are found at the edge of the morpheme which is closer to the word-edge” (Nepveu 1994:6).

‘Macro clusters’[4] break down into onset-like sub-clusters at morpheme edges. Sub-clusters are of three types: (i) a single obstruent, (ii) a single resonant or (iii) an obstruent and a sonorant, in that order. The shapes of sub-clusters and their distribution is schematised as follows:

(6)

T	T+
+R	R+
+TR	TR+

In Nepveu’s notation, T stands for any obstruent, R for any resonant, and C for any consonant.

The largest sub-cluster found in a word has the shape TR. According to Nepveu, the sub-clusters correspond exactly to the set of possible onsets allowed by the SSP, namely {T R TR}.

The question that Nepveu tries to answer concerns the nature of the sub-clusters: “What are these sub-clusters? In other words, what sort of phonological

[3] The slash in the examples demarcates where segments would violate the SSP if they were part of the following syllable.
[4] The term ‘macro cluster’ refers to the entire oversized cluster, in contrast to its constituent sub-clusters.

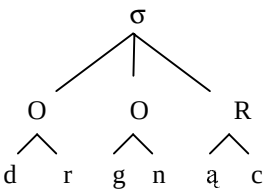
unit might they be that would account for their shape and distribution?" (Nepveu 1994:12). He considers several alternatives:

(7) a) First, they could be just what they look like, onsets and nothing more, appended to the adjacent syllables.
b) Second, they could be syllables with consonantal peaks, despite the claim that only vowels can be syllabic in Georgian.
c) The third option that is pursued by the author is that they are headless syllables, consisting of just an onset with no moraic segments.

Nepveu (1994) tries to show that the last two options are less consistent with the facts than the headless syllable hypothesis, which requires minimal additional stipulations.

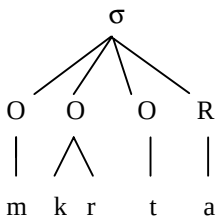
The first option was considered by Kuryłowicz (cited in Gussmann 1992) to account for oversized clusters in Polish. He proposed to analyse Polish clusters as involving the adjunction of an extra onset node to the syllable. Thus /drgnąc/ 'shiver PERF ASP' would have the following structure.

(8)



In this way all oversized onsets in Polish could be broken down into maximally two simpler onsets that are well-formed with respect to the SSP. The first problem of analysing the Georgian complex clusters in this manner is that many words would need *three* onsets to fit all the extra segments into well-formed onsets, e.g. /mkrtali/ 'pale, faint'.

(9)

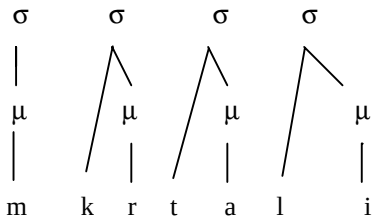


Nepveu argues that with three onset nodes in a syllable one runs the risk of having no principled way of constraining syllable size.

A second problem with extending Kuryłowicz's proposal to Georgian is that many clusters occur word-finally.[5] "As they are not followed by any syllable, the sub-clusters in these cases would have to be adjoined to the preceding syllable … Having an onset node at the end of a syllable, in the coda, simply does not make sense" (Nepveu 1994:13).

The second option against which Nepveu argues is to assume that Georgian has syllabic consonants. This approach grants independent status to the sub-clusters, and their distribution can be stated without reference to nearby syllables. Many languages have been analysed as having syllables with consonantal peaks: among these are Imdlawn Tashlhiyt Berber (Dell & Elmedlaoui 1985 and Prince & Smolensky 1993), the Mon-Khmer languages Semai, Temiar and Kammu (Shaw 1993), Polish (Piotrowski, Roca & Spencer 1992) and 85 languages reviewed by Bell (1978). An application of this approach would yield the following structure for /mkrtali/.

(10)

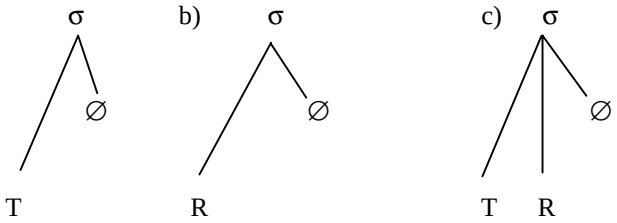


According to Nepveu (1994), the problems that arise with regard to this analysis are as follows: first, native speakers do not perceive any syllable peaks within macro-clusters; in fact, they are reported to reject the pronunciation of resonants as peaks (Nepveu cites Robins & Waterson 1952 for this claim). Moreover, stress placement is unaffected by the presence of macro clusters, which is not usually the case for languages with syllabic consonants. Additionally, the set of syllables with consonantal peaks reported for other languages is somewhat larger than the set of sub-clusters in Georgian. Finally, the /v/-metathesis facts are more mysterious than they might be if resonants act as syllable peaks in Georgian. The words /vin/ 'who' and /vardi/ 'rose' establish that /v/ can stand alone in the onset. If /r/ and /l/ were well-formed as syllable peaks, we would expect */tr.va.me.t'i/ and */k'l.va/ to be acceptable forms, but they are not. Instead, the correct forms are /tvramet'i/ 'eighteen' and /k'vla/ 'killing' (Nepveu 1994:15).

Instead, Nepveu proposes that metathesis and other facts can be explained if sub-clusters are analysed as headless syllables. His argumentation proceeds as follows: assuming that the canonical head for the members of each level of the prosodic hierarchy is a member of the next level down (Itô & Mester 1992), a headless syllable can be defined as one that contains no moras. Assuming that both nuclei and codas are moraic, this leaves only onsets as possible members of headless syllables. Headless syllables can have the following structures.

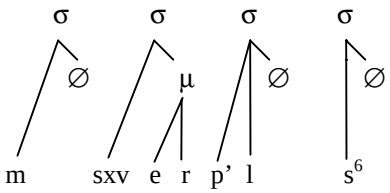
[5] Apparently, Nepveu does not make a distinction between derived and underived word structures, since consonant sequences are found in the final position of derived words and not of underived ones.

(11) a) b) c)



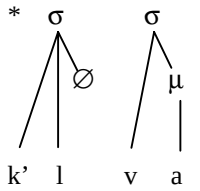
Under this proposal, /msxverp'ls/ 'victim DAT' will have the following structure:

(12)



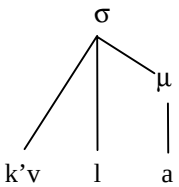
The process by which Nepveu illustrates the superiority of the Headless Syllable Analysis is /v/-metathesis. On the assumption that /l/ cannot serve as a syllable peak, it would have to be in an onset in */k'lva/, as shown below.

(13)



Nepveu says: that "... even Georgian will avoid building headless syllables if a better option is available" (Nepveu 1994:17). A better option, according to Nepveu, is available if the /v/ metathesises, allowing the formation of a single TR onset, namely /k'vl/.

(14)



Another explanation that Nepveu proposes for the metathesis is that: "… obstruents and glides are quite dissimilar, making an off-glide on an obstruent not so bad, and

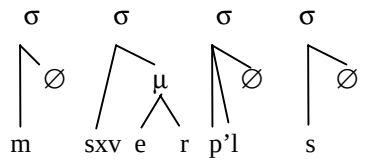
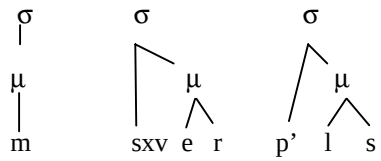
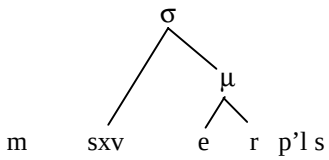
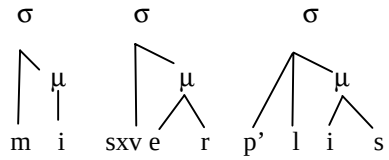
[6] The complex *sxv* in Nepeu's analysis is expressed only by a single T symbol, assuming that the *sx* is a harmonic cluster and *v* an off-glide on the primary articulation.

metathesis takes place to create the less dispreferred combination" (Nepveu 1994:11).

Nepveu concludes the discussion about metathesis as follows: "… the headless syllable analysis amounts to the null hypothesis, relying only on stipulations that would have to be made in any case. If the sub-clusters are headless syllables, the SSP will already ensure that onsets, and therefore sub-clusters, will belong to the set {T R TR}. Stress rules standardly ignore non-mora-bearing segments (i.e. onsets). Finally, headless syllables minimize the number of stipulations to account for the metathesis of /v/" (Nepveu 1994:17).

After introducing the basics of Optimality Theory, Nepveu proposes some changes in Gen. One of the sub-parts of Gen is a constraint that requires syllables to contain nuclei (Prince & Smolensky 1993:25). Nepveu calls this constrain NUC. The presence of NUC in Gen captures the common intuition that the nucleus is the one obligatory constituent in the syllable. NUC completely prohibits headless syllables; hence Gen could never produce a candidate that contains none. If the sub-clusters in Georgian are to be analysed as headless syllables, NUC must be removed from Gen and ranked with other constraints. For the Georgian data, Nepveu proposes to rank NUC below PARSE and FILL, as well as *P/[+cons], a constraint against having consonants as syllable peaks. PARSE requires that every segment in the input be dominated by prosodic structure in the output ("no deletion"; all underlying segments are phonetically realised); FILL requires that syllable positions be filled only with underlying segments ("no epenthesis"). *P/[+cons] is related to a family of constraints that Prince & Smolensky use in their analysis of Imdlawn Tashlhiyt Berber: each segment in the phonemic inventory gets a separate constraint along the lines of *P/t, which means "*t* cannot be a syllable peak". Nepveu notes that we do not need so much detail for Georgian; it will be sufficient to have a single constraint *P/[+cons] to rule out consonantal syllable peaks. The following tableau shows that NUC is crucially ranked below these three constraints in Georgian.

(15)

	*P/[+cons]	PARSE	FILL	NUC
 ☞				***
	*!*			
		*!***		
			*!*	

The pointing hand indicates the optimal candidate according to the ranking, which is the actual output; exclamation marks indicate fatal violations. Under the headless syllable analysis, NUC is violated three times in *msxverp'ls*; however, each of the other three candidates avoids violating it by violating one of the higher-ranked constraints. The second candidate has consonantal nuclei and violates *P/[+cons]. The third candidate violates PARSE by simply leaving out the problematic segments. And the last one has vocalic syllable peaks inserted, violating FILL.

According to Nepveu, this constraint ranking gives us the ability to generate the headless syllables, which are needed to account for the shape of the sub-clusters. The next step is to explain the distribution of these headless syllables.

The proposal is that headless syllables appear only at morpheme edges, and only one per edge. Moreover, even though both edges of a root could in principle host a headless syllable, Nepveu found no examples where this is the case. Having allowed for headless syllables, Nepveu tries to explain the fact that no word in Georgian is vowelless, though some words come close, e.g. /gvprckvnis/ 'he is peeling us'. Though there are many vowelless roots, no vowelless root ever exceeds the form TRTR in complexity. Such roots are always immediately followed by a

vowel in surface forms; if there is only one vowel in a word, it will be in a root or directly adjacent to it, as in /gvprckvnis/.

Each of these phenomena displays a connection between the phonological and morphological structure of the word, which brings to mind Prosodic Morphology. The basic tenet of Prosodic Morphology (McCarthy & Prince 1993) is the claim that morphological constituents are defined in terms of phonological units. This notion can help to explain the need to have a vowel in every word and the restriction of headless syllables to morpheme edges. Both conditions lead to the limitation on root complexity.

First, the requirement that every word contains a vowel is clearly a word-minimality effect, but the picture is complicated by the presence of headless syllables in Georgian. In Itô & Mester (1992), word minimality is captured by a headedness requirement on prosodic categories.

(16) *Proper Headedness*
Every (non-terminal) prosodic category of level *i* must have a head, that is, it must immediately dominate a category of level *i*–1.

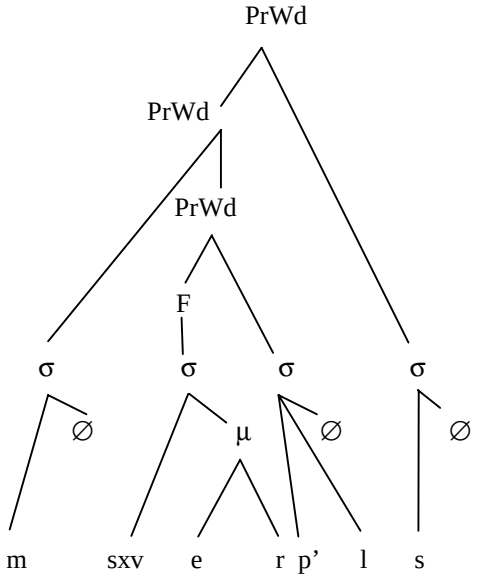
This means that every Wd must contain at least one F, every F must contain at least one σ, and every σ must contain at least one μ. This chain is broken by the introduction of headless syllables, and Nepveu tries to find a way to re-establish it. He suggests positing a high-ranking constraint on the proper headedness of prosodic words and feet that should ensure that every word contains a vowel, despite the low ranking of NUC, the constraint on the proper headedness of syllables. To ensure minimal word size, Nepveu proposes to rank foot- and prosodic word-headedness constraints (respectively HEAD (Ft) and HEAD (PrWd)) over FILL, so that they can force epenthesis. However, he cannot give examples where this happens.

Next, the restriction of one headless syllable per morpheme is related to the requirement that a morphological stem corresponds to a prosodic word, which is observed in many languages. His claim is that:

(17) "The distribution of headless syllables is not one per morpheme, but one per PrWd" (Nepveu 1994:22).

Each affix in a word is added to a stem, the minimal stem being the root. When an affix is added, a distinct stem is created, to which another affix might be added. Because each stem should correspond to a distinct prosodic word, this results in recursion of PrWd in the phonological representation of a morphologically complex word. After application of the algorithm to /msxverp'ls/ 'victim', the following structure is derived.

(18)



The syllable level and below is the same as in the structure in (12). The lowest PrWd dominates only root material; the next include the prefix /m-/, and the highest includes the suffix /-s/. It can be seen that each PrWd also incorporates one headless syllable at an edge. Headless syllables cannot be footed, so they are incorporated into the PrWd instead.

Nepveu offers several possible answers to explain why there can be only one headless syllable per PrWd: "Perhaps the incorporation of a headless syllable into a PrWd closes off the PrWd so that nothing more can be put in; in this it would clearly differ from incorporation of merely unfooted syllables, which occur in the middle of PrWd (cf. Itô & Mester 1992). A second option is to posit a high-ranking constraint like ALIGN (Ft, R, Ft, L), so that having headless syllables between feet, and in the middle of a PrWd, would be a fatal violation. A third option would be to have feet align with one edge of the word, so that headless syllables are forced to the other edge; however, this seems to run afoul of the fact that headless syllables may occur at either the right or left edges of morphemes" (Nepveu 1994:23). He is unable to provide convincing support for any of these alternatives, and makes the following stipulation.

(19) "A PrWd may contain one headless syllable, which must appear at an edge of the PrWd" (Nepveu 1994:23).

Nepveu specifies a mechanism to ensure the correspondence of stems to PrWd's with the constraints ALIGN and ALIGN-L. He chooses these constraints rather than a unitary constraint Stem = PrWd, because the left and right edges of Georgian stems

do not align equally well with PrWd's; use of the twin constraints ALIGN and ALIGN-L is better suited to capture this asymmetry, since each can be violated separately.

ALIGN and ALIGN-L will generally cause stems to correspond to PrWd's, though they are massively violated throughout the lexicon. At this point the complexity of vowelless roots and the fact that they are always followed immediately by a morpheme that begins with a vowel should be relatively transparent. Finally, he gives three conditions that are particularly relevant in this respect.

(20) a) The lowest PrWd must dominate all of the root, and must contain a vowel (HEAD(Ft)/HEAD(PrWd)).
b) Epenthesis of a vowel into a root is prohibited (FILL).
c) No more than one headless syllable is allowed per PrWd (by stipulation).

(Nepveu 1994:24)

According to Nepveu, a vowelless root of the TRTR shape is the largest that can meet all these requirements at once. A root of the shape TTRTR, however, would have to have either two headless syllables or a FILL violation, because otherwise it could not fit all the segments of the root under one PrWd.

7.1.1. Comments

After Nebieridze (1975), Nepveu's analysis was the first attempt to formalise the Georgian phonotactic patterns within the Generative phonology framework. It is a remarkable work in many respects. Some of its merits and shortcomings are considered in what follows below.

First of all, Nepveu correctly notices that the domain of the distribution of complex clusters is morphological in nature. However, there is a lot of uncertainty in his analysis about exactly which domain is relevant. Throughout his analysis (see (5), (17) and (19)), the domain where the sub-clusters are found varies between the general term "morpheme", the "prosodic word" and the "root". Establishing the distributional patterns of the consonant clusters is, in my view, the key to its proper analysis. The distribution of Georgian consonant clusters is related to the morphological constituency of a word, but identification of a morphological boundary is not an easy task. The problem with the morphological decomposition of a Georgian word is that even though it is predominantly of an agglutinative type, some inflectional affixes make it difficult to draw a clear borderline between morphemes. Nevertheless, such affixes retain phonotactic independence, e.g. it is nowadays very difficult to separate the /m-/ prefix from the stem in words like /mgeli/ 'wolf' or /mkrtali/ 'pale' (cited by Nepveu in (9) as problematic for the Extrasyllabic Consonant Analysis),[7] and if Georgian speakers are asked to decompose the word, the majority would probably say that /mgel-/ is a stem. Nonetheless, specialists know the origin of the /m/ and the unusual phonotactic pattern it gives rise to (e.g. *mg*

[7] However, the morphological status of /m-/ is recognised by Nepveu (see (19)).

violates the SSP), which can be explained by taking into account that knowledge. Thus, there are troublesome areas that require us to have a careful look at the developmental patterns of Georgian.

Complex clusters are analysed in terms of headless syllables that are the same as well-formed onsets (obeying the SSP). Why should one expect such constituents only at the edge of a morphological constituent or at the edge of a Prosodic Word? The only answer that Nepveu gives is that: "Perhaps the incorporation of a headless syllable into a PrWd closes off the PrWd so that nothing can be added …". Obviously this suggestion is quite vague and requires clarification.

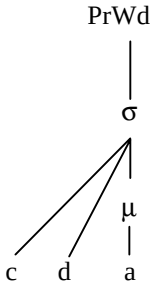
Another problem that emerges with respect to the adoption of the Headless Syllable Analysis is that after positing quite independent entities (sub-clusters) as constituents of the long consonant sequences, one can no longer capture regularities found across the whole macro clusters (see Chapter 3). Thus, the relationship that exists between the members of the consonant sequences, e.g. properties such as sharing of laryngeal specification and the decessive order of consonants, is not expressed in this type of analysis. This would be true for any type of analysis (the Semi-syllable, the Syllabic Consonant or the Extrasyllabic Consonant Analysis) that suggests a division of the long sequence without extra proposals capturing the correlations between the sub-constituents. Even though Nepveu starts his analysis by denying the arbitrariness of the composition of the Georgian consonant sequences, he is not subsequently concerned with the study of regularities of the combination of consonants in clusters, but instead proposes an analysis where there is no relation between the headless syllables (sub-constituents of consonant clusters). Thus, the arbitrariness is still maintained in his analysis.

The /v/-metathesis process is the only phonological process for which the advantage of the Headless Syllable Analysis is illustrated. Paradoxically, as Nepveu notices himself, this is the very case when headless syllables are extra, unnecessary units. Unfortunately, he thinks that any account of the process would need some stipulations, and that the fewest stipulations are needed when applying the Headless Syllable Analysis. In Butskhrikidze & van de Weijer (2001a) it was shown that /v/-metathesis can be accounted for without any stipulation, and is straightforwardly derived from observations regarding Georgian word-level phonotactics (see also 7.4.2).

7.2. Bush (1997)

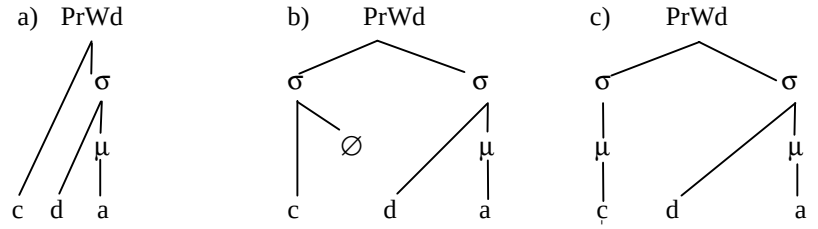
Bush (1997) argues that consonant clusters are syllabified normally: all consonants in the cluster are in the same syllable, which has a vocalic nucleus. Thus, the representation of the word /cda/ 'attempt' in the "Syllabified Consonant Analysis" (SCA), proposed by Bush, is as follows.

(21)



Bush discusses three other possible structures suggested by other authors for Georgian as well as for other languages with long consonant sequences. These representations are depicted in (22).

(22)



These representations (syllabifications) of the complex onsets correspond to three distinct analyses:

(22) a) Extrasyllabic Consonant Analysis (ECA): some consonants are extrasyllabic, attached directly to the PrWd.

b) Headless Syllable Analysis (HSA) or Semi-syllable Analysis: the empty nucleus idea originates from the theory of Government Phonology (Kaye, Lowenstamm & Vergnaud 1985; Kaye 1990). Van Lit (1988) offers an analysis of Georgian consonant sequences in which a consonant (in a sequence) is followed by an empty nucleus. The empty nucleus idea was later implemented in the 'headless syllable' analysis (Nepveu 1994) and semi-syllable analysis (Cho & King 1996): some consonants are in 'headless' syllables or 'semi-syllables', serving as onsets in syllables with an empty nucleus.

c) Syllabic Consonant Analysis (SCA): some consonants are themselves syllabic. Analysing consonants in various languages as syllabic has a long tradition (Bell 1978, Dell & Elmedlaoui 1985, 1988 and Piotrowski, Roca & Spencer 1992).

As Bush notes, these four analyses (i.e. (21) and (22a, b and c) differ in terms of the status of the SSP. In (21) the SSP is violated while in (22a–c) it is not. In (22a) the SSP is not violated because the *c* is extrasyllabic, and directly attached to the PrWd

node. The SSP inspects the sonority of members of the same syllable. In (22b) and (22c), the SSP is not violated because the *c* is in a syllable by itself, whether as an onset, as in (22b), or as a nucleus, As in (22c). In the course of the presentation of the SCA, Bush refers to the other three analyses, pointing out their strong and weak points.

Thus, according to the SCA, the SSP is not always obeyed in Georgian. Bush's account is cast in the framework of Optimality Theory, using Correspondence Theory (McCarthy & Prince 1995).

Before addressing the specifics of the SCA analysis, I list some of the assumptions adopted in Bush's analysis:

(23) a) Harmonic clusters are complex segments.
b) /v/ is not an independent segment, but a secondary articulation on the previous C. Bush represents it as an off-glide on the previous C in his transcriptions. Thus, *tkv* as in *ktven* 'you (PL)' is treated as one segment (viz. $\widehat{\text{tk}}^{\text{w}}$).
c) Affricates are treated as single segments.
d) Neither obstruents nor sonorants are syllabic in Georgian.

In his discussion of the distributional patterns of consonant sequences, Bush is uncertain about the domain. He says: "It is sometimes difficult to tell whether it is the first syllable of the morpheme or of the entire word that is relevant, but in at least most cases, it seems to be the morphological level that is the key" (Bush 1997:33). He adds: "Perhaps we need to separate out phonological prominence from some type of morphological prominence. Examples of the former case (explored in Steriade 1993) would be licensing of contrasts by peripheral syllables or stressed syllables. The latter case, applicable to Shona (Beckman 1995) and Georgian, would have to be based on something different. A possible cause would be psycholinguistic, related to the need for lexical retrieval. The idea is that the beginning of morphemes is more relevant than other parts for retrieval, and so it is a better place to locate contrast. More contrast means that more morphemes can be easily distinguished from each other" (Bush 1997:33). Later in his analysis, Bush notes that the constituency of affixes is not so complex as that of roots. He tries to explain this distinction on a functional basis. In the process of grammaticalisation, semantic reduction proceeds in parallel with phonological reduction (loss of segments, tone, stress, etc.) (Bybee, Perkins & Pagliuca 1994). However, he does not stipulate separate constraints on affixes, but accepts the ranking offered by McCarthy & Prince (1995): MAX_{root} >> $\text{MAX}_{\text{affix.}}$ This is probably the reason why the general term 'morpheme' is used in the rest of his paper in the generalisations on the distributional characteristics of the consonant sequences, for instance in (24).

(24) a) Only the first syllable of a morpheme allows SSP-violating clusters.
b) The second syllable of a morpheme allows only clusters that do not violate the SSP.
c) The third syllable of a morpheme does not allow any clusters at all (and there are no longer native Georgian morphemes).

(Bush 1997:14)

As shown in (24), these distributional regularities regarding consonant sequences are formulated in terms of the syllable, specifically by counting syllables in a word. This is to say that phonotactic patterns can be accounted for by reference to the syllable. Moreover, Bush argues that there is other phonological and phonetic evidence for Georgian making reference to the syllable, which supports the claim of consonant sequences being syllabified together in a syllable. The phonological and phonetic evidence discussed by Bush concerns first the Monosyllabic Lengthening (ML) rule. The rule concerns the fact that in yes-no questions, monosyllabic words that are pronounced with a sharp rise in intonation are lengthened into two syllables. To account for this phenomenon, a distinction between words with one syllable and words with more than one syllable is needed. As Bush suggests, this phenomenon provides further support for an analysis in which consonant sequences are syllabified normally. "At the very least, it shows yet again that any putative 'headless syllables' are not visible to the syllable-counting mechanism relevant to ML" (Bush 1997:29). Bush illustrates the process with words which have initial clusters and only one vowel.

(25) a) k'artópils prckvniis/*prckvnis
potatoes (he) peels
'Is he peeling the potatoes?'
b) sač'iroa cdaa/*cda?
necessary (is) attempt?
'Is an attempt necessary?' (Bush 1997:28)

As Bush notes: "The Monosyllabic Lengthening facts argue against the Headless Syllable Analysis and Syllabic Consonant Analysis, but say nothing about the Extrasyllabic Consonant Analysis. This is because the HSA and SCA posit syllables that turn out not to be counted for the purposes of ML, while the ECA posits no extra syllables. Since most consonants in clusters would be simply extrasyllabic, under that analysis we would have no reason to expect that there would be any extra syllables in words with consonant clusters" (Bush 1997:31).

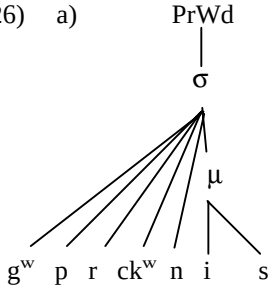
Bush adopts the idea that syllable structure is coextensive with timing relations between segments (Browman & Goldstein 1989). The correlation is illustrated by way of the phenomenon of word-initial devoicing and a C-centre experiment.

Considering the voicing of the following consonant, Bush finds that nasals are significantly more likely to devoice before a voiceless consonant than before a voiced consonant. That is, words like /mdgomareoba/ 'situation' are less likely to have a voiceless [m] than words like /mta/ 'mountain' are. Devoicing does not occur

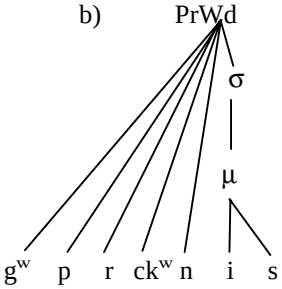
word-internally before voiceless consonants in words like /šu.am.ta/. There is no devoicing in /šu.am.ta/ (proper name) because the syllabifications and therefore the timing relations are different. In /šu.am.ta/, the /m/ and /t/ are in separate syllables. "Since syllabification is coextensive with timing relations, the syllable boundary between /m/ and /t/ means that there is a greater time separating them than there would have been, were they syllabified together" (Bush 1997:21–22).

Based on Browman & Goldstein (1988) and Fowler & Tassinary (1981), Bush designed an experiment that should yield another phonetic argument for the syllabification of consonant sequences. Browman & Goldstein (1988) introduced the notion of *C-centre*, a point roughly in the middle of a sequence of consonants, which is in a fixed temporal relation with the onset of the vowel. More formally, "[the C-center] is a simple global property of this sequence that was derived as follows: for every consonantal gesture, the (temporal) midpoint between the left and right edges of the plateau was computed. The C-center of a sequence is the mean of all the midpoints of the gestures in that sequence" (Browman & Goldstein 1988: 144). The more consonants in the sequence, the earlier the C-centre will be with respect to the acoustic onset of the vowel. But not merely any consonants in a row will count: this concept depends on the syllable being the unit of segmental organisation, so that it only applies to consonants syllabified together. Bush considers two 'basic' analyses of /gvprckvnis/ 'he is peeling us':

(26) a)



b)



In (26a), all consonants in the cluster are syllabified together, while in (26b) they are not. The latter stands for any analysis in which the consonants are not syllabified together, whether they are extrasyllabic, in headless syllables or syllabic. If the syllable were the unit of segmental organisation, it would only be in (26a) that all of these consonants in clusters would be organised in some way with respect to the syllable. In (26b), only one segment will be organised with respect to the syllable. Thus, if there is evidence that all the consonants in a cluster are counted together in determining the C-centre, Bush has evidence that they are syllabified together.

In fact, Bush did not run any experiments on the C-centre but instead on the perceptual centre (p-centre) proposed by Fowler & Tassinary (1981) and Fowler (1983). As summarised by Browman & Goldstein, "when asked to align a list of words with different initial consonants in a regular rhythm, listeners do not align the

successive words at their acoustic onsets". Instead, they line the words up with respect to a p-centre, which bears a close resemblance to the articulatory C-centre.

Bush's experiment was conducted on one subject. He first devised 33 utterances, each consisting of seven nonsense words. Nine were homogeneous utterances (i.e. the same syllable was repeated seven times), one for each of the syllables *is, tis, dis, gis, mis, nis, ris, bis, pis.* The remaining 24 were alternating utterances, using the pairs *tis*/*is, ris*/*is, gis*/*is, nis*/*is, tis*/*is, mtis*/*tis, rtis*/*tis, tris*/*ris, sgis*/*gis, cdis*/*dis* and *prckvnis*/*gvprckvnis*; two utterances were constructed from each pair, one with the first word of the utterance being the first in the pair, and one with the first word of the utterance being the second in the pair. Bush made four copies of each homogeneous utterance and two of each alternating one, and presented them to an informant. He instructed the informant to pronounce each utterance "as rhythmically as possible", first without and then with the aid of a metronome.

The preliminary results that Bush obtained are as follows. First of all, words like *tis* and *ris* were lined up with the metronome beats very close to the start of the first consonant. The release of the *t* in *tis* was almost exactly at the metronome beat, while the *r* in *ris* usually started about 100 ms before the beat and continued for about the same amount of time after it, followed by the acoustic onset of the vowel. Let us now concentrate on the pair *rtis*/*tis.* Since the metronome beat is earlier (with respect to the acoustic onset of the vowel) in the first word, this means that the p-centre is earlier in the first word than in the second. This, in turn, indicates a shift in the C-centre (if the p-centre is indeed the same as the C-centre). And, since the shift in the C-centre can only be caused by the addition of a consonant that is syllabified together with the other consonants in the sequence, it can be concluded that *r* in *rtis* is indeed syllabified with *tis*. In the case of the pair *prckvnis*/*gvprckvnis*, for *gvprckvnis* the beat comes about 520ms after the start of the first consonant. Interestingly, the beat comes right before the *n.* As Bush notes, the beat is in fact earlier with respect to the acoustic onset of the vowel in *gvprckvnis* than in *prckvnis*, as he would predict to be the case if these consonants are syllabified together.

Thus, the C-centre experiment shows that the consonants in clusters act as if they are in a syllable together, because they seem to be organised as a group in relation to the nucleus.

Bush concludes from the phonological processes and phonetic facts that all consonants in complex clusters are syllabified together in a syllable with a vocalic nucleus. This means that many consonant clusters will violate the Sonority Sequencing Principle. But this does not mean that under this analysis the SSP must be given up entirely. The SSP appears to be violated only by clusters which occur in the first syllables of morphemes. To capture the generalisation within Optimality Theory he proposes to split Faithfulness in two: $\text{MAX}_{\sigma 1}$, which only applies to the first syllables of morphemes, and ordinary MAX, which militates against underparsing of segments. $\text{MAX}_{\sigma 1}$ inherently dominates MAX, while SSP is ranked between them: $\text{MAX}_{\sigma 1}$ >> SSP >> MAX. The constraint $\text{MAX}_{\sigma 1}$ is defined as follows:

(27) MAX$_{\sigma 1}$: For each segment in the root-initial syllable of the Input, there must be a corresponding segment in the Output.

'The first syllable of the morpheme' is not determined according to its final output syllabification, but according to some 'preliminary' syllabification of just the morpheme (Bush 1997:34). The output of this level, consisting of syllabified morphemes, is input for the final level. The syllabification from the preliminary level would determine what counts as the first syllable of the morpheme for MAX$_{\sigma 1}$ in deriving the final form. Thus, even though levels are not widely accepted in OT,[8] Bush needs to introduce them in his analysis.

To ensure that segments get parsed into syllables, Bush furthermore introduces the constraint PARSE (SEG → σ).

(28) PARSE (SEG → σ): every segment must be parsed into a syllable.

This constraint must be obeyed even when, as is often the case, a morpheme has no vowels. He argues that there are no vowelless syllables at the output level. To let vowelless morphemes be parsed into syllables, he allows consonantal nuclei at the level of preliminary syllabification, formalised by the constraint *P/C. *P/C militates against consonants being parsed as syllable peaks. To ensure that no segments are deleted at this preliminary level, he includes MAX as an undominated constraint. The effect of these three constraints is shown in (29).

(29)

/pr͟c͟k^{w}n/	PARSE (SEG → σ)	MAX	*P/C
☞pr͟c͟k^{w}n̩			*
.pr.c͟k^{w}n̩	*!*		*
.p͟r̩.c͟k^{w}n.			**!
<pr>.ckvn̩.		*!*	*

The first candidate is a winner because it violates only the lowest-ranked constraint *P/C. The second candidate, which has two segments unassociated to a syllable, violates the highest-ranked constraint PARSE (SEG→σ). The third candidate has two syllabic consonants and thus violates *P/C twice, and the fourth candidate, with two deleted consonants, violates the constraint MAX, which militates against deletion of any segment at the preliminary level.

Another constraint that is relevant to preliminary syllabification is ONSET MAXIMIZATION. The constraint comes into play when hypothetical inputs with word-medial clusters are considered.

(30) MAXONS: In the configuration Nuc X Nuc, X contains no complex coda.

[8] See Booij (1997), who argues for the reintroduction of levels into OT. McCarthy & Prince (1993), in their analysis of Axininca Campa, also make use of a 'prefix' vs. 'suffix' level.

At this level the SSP must be dominated by all four constraints: PARSE (SEG → σ), MAX, MAXONS and *P/C.

Let us take the root /targmn-/ 'translate' to illustrate the preliminary syllabification suggested by Bush.

(31)

/targmn/	PARSE (SEG→σ)	MAX	MAXONS	*P/C	SSP
☞.targmn.					*
.tar.gmn̩.				*!	*
.tar.<gmn>		*!**			
.tar.gmn	*!**				
.targ.mn̩.			*!	*	*

The first candidate is a winner because it violates only the lowest-ranked constraint SSP. Each of the other candidates violates one or more of the four higher-ranked constraints.

This gives the final ranking for the level of preliminary syllabification.

(32) PARSE (SEG→σ), MAX, MAXONS >> *P/C >> SSP

The output of the preliminary level serves as the input to the final level, where it will be resyllabified and subject to the rest of the constraints of Georgian. In the case of /targmna/ 'to translate', after the suffix /-a/ is added to the root /targmn-/ the ranking will be as follows.

(33)

.targmn̩., .a.	$MAX_{\sigma 1}$	MAXONS	SSP	$MAX_{\sigma 2}$	*COMPLEX	MAX
☞.tar.gmna.			*		*	
.targ.mna.		*!	*		**	
.tar.<gm>na.	*!*				*	**
.targmn.na.		*!	*		*	**

Since consonants are not licit syllable peaks in the final syllabification of Georgian words, *P/C must be undominated at the final level. This ensures that there will be no syllabic consonants in Georgian, and thus prevents structures allowed by the Syllabic Consonant Analysis from surfacing. Similarly, to ensure that no consonants surface as extrasyllabic (as allowed by the Extrasyllabic Consonant Analysis), PARSE (SEG→σ) is also undominated at the final level.

To summarise, in his OT analysis of the Georgian facts, Bush distinguishes two types of constraint ranking: one for the preliminary syllabification level (see (34)), and another for the final level (see (35)).

(34) PARSE (SEG→σ), MAX, MAXONS >> *P/C >> SSP

(35) $\text{MAX}_{\sigma 1,}$ MAXONS >> SSP >> $\text{MAX}_{\sigma 2}$ >> *COMPLEX >> MAX >> NOCODA, ONS

Since loan words and foreign names do not follow the generalisation that consonant clusters do not occur in the third syllable of morphemes. Bush suggests treating them separately from native Georgian words. Problematic words are:

(36)

or.to.gra.pi.a	'orthography'
dos.to.ev.ski	'Dostojevsky'
kro.no.me.tri	'chronometer'

For these words Bush proposes to rerank constraints as suggested in their account of the periphery of the lexicon by Itô & Mester (1995).[9] Thus, the constraint ranking for the word *ortograpia* will be as follows.

(37)

/ortograpia/	$\text{MAX}_{\sigma 1}$	$\text{MAX}_{\sigma 2}$	MAX	SSP	*COMPLEX
☞.or.to.gra.pi.a.					*
.or.to.g.ra.pi.a.			*!		
.or.to.g.<r>a.pi.a.			*!		
.or.to.<gr>a.pi.a.			*!		

In a statistical account of the complex clusters, Bush finds an interesting generalisation: "the fewer the syllables in a word, the longer the consonant clusters tend to be; the more syllables, the shorter the clusters. This generalisation seems to reflect an interesting principle at work: there is an interesting trade-off between the number of syllables and the length of consonant clusters" (Bush 1997:49). He does not explore the principle further. The correlation is discussed in Chapters 1 and 6.

7.2.1. Comments

Compared to Nepveu's analysis, Bush's analysis is based on a comprehensive range of arguments, involving phonological processes, phonetic experiments, statistical data, etc. Thus, his work represents significant progress in the understanding of Georgian word-level phonotactics. However, there are some shortcomings that I would like to mention.

An empirical problem with the analysis of /cda/ 'attempt, to try' in (11) is that the *cd* cluster cannot be analysed as a true cluster, since there are forms with the same root that have a vowel between the consonants, e.g. /e-cad-e/ 'try you', /v-e-cad-e/ 'I tried'. The word /cda/ is in fact polysemous in Georgian. Another meaning of /cda/ is 'to wait'. Also, in this case a form exists with the vowel /u/ between the

[9] However, Rhee (2002) argues that loan-word phonology is not completely different from the native phonology in Korean, e.g. the behaviour of word-internal epenthetic vowels in native words in Korean is the same as that in loan words.

consonants, e.g. /v-u-cad-e/ ‘I waited for him’. Thus, the /cda/ form is derived from the stem /cad/, in which the vowel /a/ is deleted between the consonants /c/ and /d/. Another observation with respect to the same form is that Bush considers the *cd* cluster to be an instance of a violation of the SSP. The consonant /c/ is a voiceless affricate in Georgian, and the consonant /d/ is a voiced stop. It seems therefore that Bush is following a sonority ranking in which all affricates are more sonorous than stops, regardless of their laryngeal specification. However, I know of no arguments for regarding affricates as different from stops in terms of sonority, either in Georgian or in other languages. The cluster *cd* is considered well-formed in the Gradual Consonant Analysis, assuming the Compensatory Principle proposed by Melikishvili (1997) (see Chapter 2).

Another problem with the SCA is empirical as well as theoretical. As Bush notes himself: “Without the SSP, we expect to have clusters of arbitrary contours and length, like *rmqb*, *tdtdtd*, or *trblqsxvntkb*. While it is true that only a small percentage of the clusters that would arise from random combinations of consonants are actually attested, I claim that this fact does not need to be accounted for by the phonology” (Bush 1997:48). This would mean that the fact that only the first syllable allows arbitrary combinations of consonants should not be of any concern to phonological theory. This seems to be the reason why Bush does not discuss the composition of long consonant sequences at all in his analysis. From the very beginning, he abandons any attempt to find out what these clusters are constituted of, and concentrates only on their distributional patterns. I disagree with this strategy, and consequently also with the statement cited above. The claim that the complex clusters in the first syllable of a word are arbitrary is not empirically justified (see Chapter 3). Consonants occurring in sequences in Georgian obey certain distributional regularities, and phonological theory should try to account for these.

At various points, Bush’s arguments for the Syllabified Consonant Analysis do not seem to be very consistent. On the basis of evidence from the phonological process of Monosyllabic Lengthening and the P-centre experiment, he draws the conclusions that all consonants should be syllabified together with one nucleus, but later in his analysis he says that, at the level of ‘preliminary syllabification’: “To let vowelless morphemes be parsed into syllables, we must allow consonantal nuclei at the level of preliminary syllabification” (Bush 1997:35) (see also (29), where syllabic *n* is accepted in the analysis). Thus, different levels allow conflicting syllabification patterns, which leads to arbitrariness in the analysis.

The facts of Monosyllabic Lengthening form a good argument, in general, to say that consonants are not syllabic. However, if they are in a constituent which does not contribute to weight, this argument does not help. As for the argument from word-initial devoicing, while it is true that any assimilation process shows that segments that are assimilated are closely connected, this does not prove that the segments should be necessarily in one constituent (i.e. in one syllable). From cross-linguistic studies, and in Georgian as well, it emerges that many assimilation and dissimilation processes are attested between segments that neither belong to one syllable nor are monomorphemic e.g. /mi-s.-dev-s/ > /mi-z.-dev-s/ ‘runs after’. The con-

sonants /s/ and /d/ are neither in the same syllable nor in the same morpheme; regardless of these conditions, the assimilation process, i.e. *sd* > *zd*, does occur.

A brief comment is in order with regard to the P-centre experiment and its validity for the issue of the syllabification of consonants in Georgian. Bush (1997) builds up the P-centre experiment on the basis of the tentative proposal made by Browman & Goldstein (1988), in which the perceptual P-centre bears a close qualitative resemblance to the articulatory C-centre. The suggested correlation, however, has not yet been proven. As Goedemans (1998) points out: "… much about the relationship between P- and C-centers, and even about the true nature of these points themselves, remains unclear" (Goedemans 1998:99). Putting this issue aside, the P-centre experiment on Georgian data was run on one subject only, which makes the results of the experiment inconclusive at best. In addition, the two structures given in (26) are also inconclusive, since the addition of any consonant, not only in the onset but also in the coda, moves the P-centre, and the introduction of some parameter would be necessary to evaluate the validity of the change. From the description of the experiment, it seems that this has not been done. Taking all of these aspects into account, one can conclude that Bush's P-centre experiment is inconclusive with respect to the syllabification of the consonants in Georgian. Bush himself points out that the results of the experiment are very preliminary. Much work still has to be done in the future to use such experiments as plausible evidence for making claims about structural properties of a language.

There is much uncertainty about the distributional regularities of consonant clusters as well. "It seems that the relevant generalization is not that clusters can appear at the start of words, but that they can appear at the start of morphemes. We have also seen that clusters can occur at the end of some morphemes" (Bush 1997:11). Since Georgian has vowelless roots, he needs to introduce levels as well as to posit consonantal nuclei for the 'preliminary syllabification'. If the analysis used the stem and not the root as a relevant domain to state the distributional regularities of complex clusters, it would need to posit neither levels nor 'consonantal nuclei'. This option is considered in my analysis, the GCA, in Chapter 6.

Bush considers loan words as problematic, but the words he refers to are not problematic (see (36)). One is a proper name /dos.to.ev.ski/. Cross-linguistically, proper names are special in many respects, and there are numerous Georgian proper names with a much more complex structure, e.g. /me.č'ur.č'le.tu.xu.ce.si/, with a complex cluster in the third syllable. As for /or.to.gra.pi.a/ and /kro.no.me.tri/, they are compounds (morphologically complex words). There are many native Georgian compounds with consonant sequences in the third or even the fourth syllable of a word, e.g. /ma.ma.š vi.lu.ri/ 'parentally', /sa.k'a.co.bri.o/ 'worldly', etc. These examples again suggest that the stem is the relevant domain. Two stems in compounds would have consonant sequences twice, as expected, and the examples cited above would not have to be treated separately.

Bush's statistical study of the correlation between the length of a consonant cluster and the number of syllables in a root is very interesting and it is true that long consonant clusters are predominantly found in monosyllabic or disyllabic roots. Un-

fortunately, Bush does not explore the generalisation further, and does not explain or implement it in his analysis.

Bush virtually disregards the asymmetry between lexical and grammatical morphemes, and accounts for the distributional patterns of consonant sequences only in terms of counting syllables. This type of strategy later raises problems as to why violation of the SSP in the first syllable can occur, while clusters are found without SSP violation in the second syllable and no clusters are found in the third syllable. There is no explanation for these patterns and there is no way to explain them without referring to the asymmetry between lexical and grammatical morphemes. The structure proposed by Nepveu in (18) quite explicitly demonstrates morphological composition and relates it to phonological complexity in terms of headless syllables. For more about this correlation see also Chapter 6.

7.3. Toft (1999)

Toft's (1999) paper "Grunts and gutturals in Georgian (an investigation into initial consonant clusters)" aims to investigate whether sequences of consonants in Georgian constitute true clusters in Georgian phonology or only apparent ones. This paper is cast in the Government Phonology framework. Theory-internal issues would involve establishing (a) which, if any, constituents branch in Georgian, (b) whether empty nuclei are found in Georgian and if so (c) what conditions there are on *p-licensing*. Toft solely focuses on clusters with up to three members occurring within a monomorphemic context. Before addressing specifics of the theory of Government Phonology, in general, and its application to the Georgian data in particular, I introduce some assumptions adopted by Toft with respect to the phonotactic patterns of Georgian.

(38) a) Harmonic clusters are not considered as complex segments (on the basis of the acoustic study of Chitoran 1998).
b) There are no long vowels or diphthongs in Georgian.
c) Toft posits schwa /ə/ alongside the five vowels of Georgian. This schwa is found only within (orthographic) consonant sequences and nowhere else.
d) Toft notices some variability of the /v/. She cites two contexts: (i) Stem-final -/v/ is optionally pronounced as /w/ when followed by the nominative suffix, e.g. [tagwi] for /tagvi/ 'mouse' (Toft 1999:286) and (ii) Stem-final /v/ is optionally lost when followed by the ergative suffix e.g. "pilt'uma 'lung' + ergative suffix < *pilt'v-*" (Toft 1999:286). Unlike Nepveu (1994) and Bush (1997), Toft's analysis does not assume the defective nature of /v/.

The version of the Government Phonology (henceforth GP) adopted by Toft is one utilising the Revised Theory of Elements, where (a) the concept of 'charm', which

restricts the combination of elements has been replaced by Licensing Constraints, (b) the number of elements has been reduced to A, I, U, L, H, ? and (c) where all constituents may branch. The internal representations of Georgian consonants as proposed by Toft are as follows:

(39)

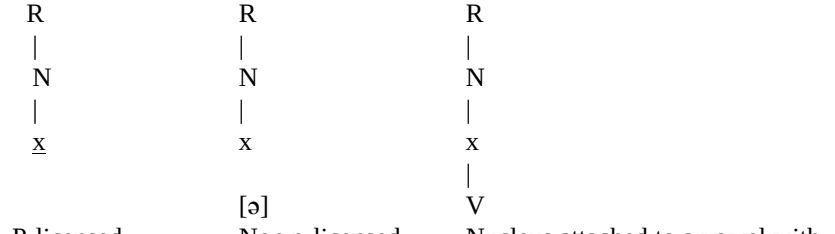
b (U.?.L)	p (U.?.H)	p' (U.H.?)
d (A.?.L)	t (A.?.H)	t' (A.H.?)
g (?.L)	k (?.H)	k' (H.?)
ǰ (I.?.L)	č (I.?.H)	č' (I.H.?)
j (A.I.?.L)	c (A.I.?.H)	c' (A.I.H.?)
		χ' (?)
z (A.L)	s (A.H.)	
ž (I.L)	š (I.H)	
ɣ (?.L)	x (H)	
l (A.?)		
r (A)		
m (U.?.L)		
n (A.?.L)		
v (U.H) and (U) (/v/ and /w/)		
h (H)		

Toft proposes the following constraints on government in Georgian.

(40)
a) Governors must be headed.
b) Governees must be headless.
c) Within a governing relationship the governor and governee may have no more than one element in common (e.g. the sequences */bn/ and */gm/, which share two elements, should not be found).

Incorporating the assumptions (40a) and (40b), Toft proposes three possible structures for nuclei in Georgian.

(41)



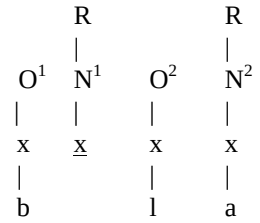
P-licensed empty nucleus	Non p-licensed empty nucleus	Nucleus attached to a vowel with 'genuine' melodic content

In her investigation of the nature of Georgian onsets, Toft discusses the following two-term sequences.

(42)		*Transliteration*	*Transcription*	
	a)	blagvi	blagvi	‘blunt’
		braldebiti	braldɛbiti	‘accusative’
		grili	grili	‘cool’
		lboba	ləboba	‘softening’
		rbili	rəbili	‘soft’
		rgoli	rəgoli	‘link’
	b)	bgera	bəgɛra	‘sound’
		gdeba	gədɛba	‘throwing away’
		tkeši	təkɛši	‘downpour’
		tbili	təbili	‘warm’
	c)	bneli	bəneli	‘dark’
		gmiri	gəmiri	‘hero’

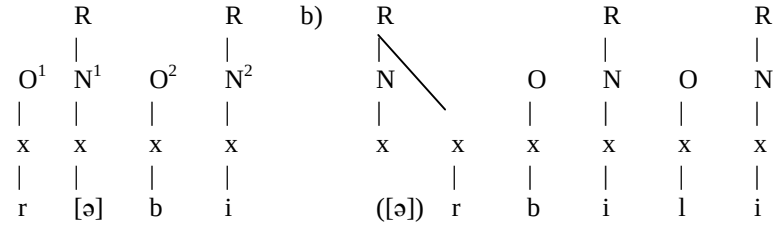
On the basis of the first three examples in (42a), Toft assumes that Georgian may have branching onsets. However, the other three examples in (42a), and the ones in (42b) and (42c), do not show branching structures according to Toft. Regardless of this distinction, she argues that the clusters *bl*, *br* and *gr* can also be analysed as spurious. Thus, the analysis for /blagvi/ is as follows.

(43)



As for the clusters with the reverse order of consonants, two structures are posited. For example, for the word /rbili/ Toft says: “... both rəbili and (ə)rbili are documented” (Toft 1999:287). Thus, two structures are proposed:

(44) a) b)

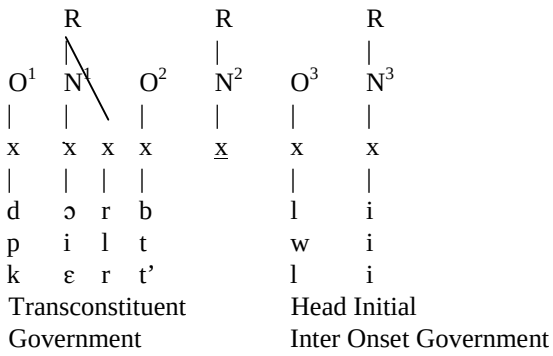


The structure in (44b) seems plausible for Toft, because branching rhymes are proposed in Georgian.

Based on the data presented in (42), Toft argues that: "Head Initial Inter Onset Government is a condition on the p-licensing of empty nuclei in Georgian and the first nucleus in Georgian is accessible to p-licensing" (Toft 1999:284). Toft proposes that the p-licensed empty nuclei may act as government licensors. Consider the words /dorbli/ [dɔrbli] 'dribble', /pilt'vi/ [piltwi] 'lung', /kert'li/ [kɛrt'li] 'dandruff'.

(45) *Direct Government*

Licensing Proper Government



Toft gives the following explanation for the structure depicted in (42): "N^2 here gives license to govern the segment to its left (Transconstituent Government). The question arises as to what O^2 gives the license to govern O^3 (Inter Onset Government). It may be that N^2 licenses O^2 to govern O^3" Toft (1999:290).

To summarise, the proposals made by Toft are as follows.

(46) a) Nuclei do not branch.
b) Onsets do not branch.
c) Rhymes may branch.
d) Head Initial Inter Onset Government is a condition on p-licensing of properly governed empty nuclei.

7.3.1. Comments

There are some empirical as well as theoretical problems with Toft's analysis. I will start with the empirical ones.

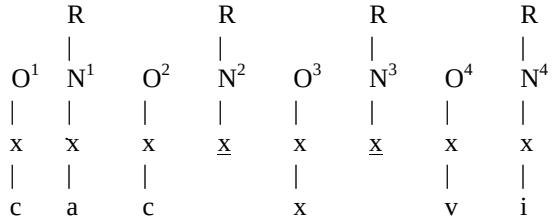
Toft's study of the consonant clusters of Georgian is limited to a certain type of sequence (of up to three consonants) as well as to the position in which they occur (word-initially, although other positions are also referred to in the course of the argumentation). This should be the reason why the nature and behaviour of some sequences of the consonants, as well as the specifics of some phonemes are not properly appreciated. These remarks concern both the status of harmonic clusters and the sonorant /v/.

Both perceptual studies (see Chapter 5) and the phonological behaviour of harmonic clusters (see Chapter 3) support the treatment of harmonic clusters as complex segments, for which Toft (1999) does not present counterevidence. Note that it would make her job easier if she recognised complex segments, because fewer empty nuclei would have to be postulated. Thus, she would avoid introduction of extra structural constituents.

Toft mentions the variability of the sonorant /v/ (see (38d)), but the contexts she cites are not relevant for the realisation of the specific allophone of /v/. The sonorant /v/ certainly has the allophone [w] in the word /tagv-i/, but not because it is followed by the nominative case marker, but rather because it follows a consonant, the velar /g/. The same applies to the discussion about /v/-loss in (38d) (more about the process can be found in Chapter 3). Finally, the defective status of /v/ is commonly recognised by almost all authors working on Georgian phonology, but remains unappreciated in Toft's analysis.

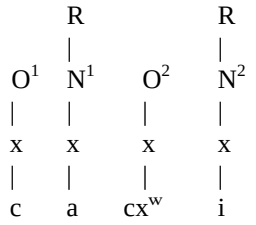
Without accepting the treatment of harmonic clusters as complex segments and the defective nature of /v/, Toft would have to analyse the word /cacxvi/ 'lime tree' as follows.

(47)



Adoption of the two proposals would yield a much simpler representation.

(48)



The representation in (48) is simpler by two onset nodes, one rhymal node and one rhymal-complement node, and is obviously a more elegant structure than the one depicted in (47).

Another point that needs to be discussed is that Toft posits schwa in the phonological inventory of Georgian, which is not substantiated by any phonetic or phonological studies. Native speaker intuitions also do not support the introduction of schwa in Georgian. It is proposed that schwa appears not only between consonants of any onset cluster, but before them as well, as in the case of /(ə)rbili/ 'soft'. Toft

says that this form is documented, but does not refer to the source. Besides, there are no limitations on how many schwas can occur in a word. The loan-word phonology also does not provide any evidence for positing schwa.

As for the empty constituents, there are several methodological ways to postulate them in the representation. One of these, mentioned by Toft at the beginning of the paper, is the comparative method: "Following a comparison of cognates in the various Kartvelian languages, Žgent'i argues that some of the consonant clusters in Georgian are the result of vowel reduction over time, cf. Georgian /jma/ 'brother', Megrelian /ǰima/, Svan /ǰemil/. This is one indication that some consonant sequences in Georgian may in fact be spurious clusters; orthographic collocation of consonants is not an adequate basis for their classification as clusters" (Toft 1999:277). I agree with this methodological strategy and actually apply it widely in my analysis. Unfortunately, Toft does not pursue this strategy throughout her subsequent analysis. All empty constituents in her analysis are posited for theory-internal reasons. Synchronic vowel ~ zero alternations in paradigmatically related forms are not taken into account.

One of the mechanisms employed for generating empty constituents in Government Phonology is the Licensing Principle, which justifies the existence of a certain unit in a structure because of some prosodic or phoneme internal constituency properties. Moreover, one licensed or non-licensed unit may license another unit. It would take me too far afield to evaluate the licensing proposals made by Toft in this thesis (see Rhee 2002 for a Government Phonology account of empty nuclei in Korean).

Regardless of any methodological disagreements, Toft arrives at a conclusion which I share with her. The formulation proposed in this thesis is more general and says:

(49) All surface clusters are the result of vowel deletion.

However, different ways are chosen by Toft and the present author to arrive at this conclusion. A priori, Toft posits schwa between consonants in onset clusters, while I take into account paradigmatically related forms and the complexity of some consonant sequences (harmonic clusters and C + /v/ combinations), and also apply the comparative method. The latter strategy, as will be demonstrated in the following section, is explanatorily and empirically more adequate.

Additionally, Toft claims that all Modern Georgian clusters are spurious, while I argue that in Modern Georgian consonant sequences are maximally biconsonantal. The claim is important in accounting for the metathesis facts (see the following section).

7.4. Discussion and conclusions

The goal of this section is to evaluate and demonstrate which of the four types of analyses discussed in this thesis is superior. In order to address these issues, this section will be divided into two subsections: evaluation and demonstration. The evaluation part involves a scrutiny of each approach against certain criteria spelled out below. The demonstration part involves considering a certain phonological process in terms of each of the four types of analysis.

7.4.1. Evaluation

Any scientific enterprise involves three aspects: empirical, formal and conceptual (or interpretative). The task of any theory is to make the empirical phenomena intelligible, and in order to do this conceptual elements are used. Conceptual elements are theory-specific, as any theory employs its own characteristic abstractions, terminology and techniques of interpretation and explanation. The outcome of this enterprise is to arrive at a deeper understanding of the phenomena. Note that, at this point, I will not engage in a discussion about the primacy of empirical as opposed to theoretical. What will be discussed here is merely the nature and characteristics of the interrelationship between empirical and theoretical.

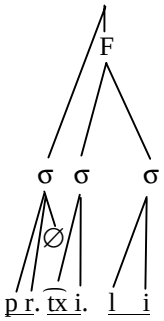
As mentioned above, there are many different ways to describe the relation between empirical facts and theoretical constructs, but there are also criteria, most of which are commonly acknowledged, which favour one description over the other. Some criteria considered in the evaluation of competing analyses are as follows.

(50) a) Simplicity or economy
b) Generality or abstractness
c) Adequacy (empirical and explanatory)

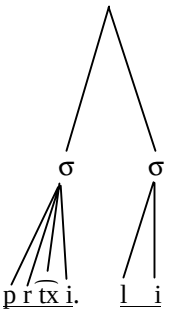
The criteria of simplicity and economy are meant to reveal how concisely (without leaving valuable information unaccounted for) and non-speculatively a theory can deal with data. The criteria of generality or abstractness are meant to establish whether the tools and terms of a theory are sufficient for cross-linguistic applicability. And finally, the criteria of empirical and explanatory adequacy establish whether a theory is empirically grounded, whether it provides with testable link between the explanandum and explanans and so has predictive power.

Taking into consideration these three types of criteria, let us consider four possible types of analysis of one word, /prtxili/ ‘careful’: (i) The Headless Syllable Analysis (Nepveu 1994), (ii) The Syllabified Consonant Analysis (Bush 1997), (iii) The analysis in Government Phonology (Toft 1999) and (iv) The Gradual Consonant Analysis (this thesis).

(51) a) PrWd

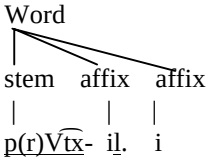


b) PrWd



c) ON ONONONON
p ə r ə t ə x i l i

d)



The relations considered in the evaluation process are threefold: (i) the relation between the word /prtxili/ and its representation, e.g. *pr.txi.li* (see (51a)), (ii) the relation between the representation, e.g. *pr.txi.li* and its interpretation in terms of certain theoretical constructs (in this case in terms of the syllable, the foot and the prosodic word) and (iii) the relation between the representation, e.g. *pr.txi.li* and the word /prtxili/. The first relation is a theory-external one, and can be evaluated in terms of empirical adequacy, simplicity and economy. The second relation is a theory-internal one, and can be evaluated in terms of abstractness and simplicity. In addition, the representations should meet theory-specific constraints and principles. The third relation combines theory-internal and external aspects and can be evaluated in terms of explanatory adequacy, simplicity and generality. This relation is supposed to demonstrate the predictive and explanatory power of a theory. It is important to note that, strictly speaking, there are no sharp boundaries between the relations which I have distinguished, and therefore they should be used merely as a guiding procedure. The overall evaluation of the analyses should establish which is superior. All four types of analyses presented in (51) are considered in turn.

The Headless Syllable Analysis (Nepveu 1994) (see (51a)) offers the representation *pr.txi.li* for the word /prtxili/. On the segmental level, the representation is empirically adequate: no segment is deleted and no segment is inserted. It is simple, but arguably not economical, because in order to analyse the long consonant sequence it introduces an extra constituent, the headless syllable. The introduction of headless syllables also posits problems for theory-internal reasons. Headless syllables are directly associated to the prosodic word, which is a violation of the Licensing Principle. The HSA is therefore abstract, but not simple. Finally, it is not clear how the representation *pr.txi.li* is related to the word /prtxili/. One cannot predict

when and why constructions like /prtxili/ will emerge in a language. In the definition of the headless syllable, there is nothing intrinsic to these constituents which would motivate the emergence of words with some long consonant sequences and not others. Thus, the strategy employed in the HSA can be regarded as one of convenience, adopted to simplify handling empirical data, which reveals nothing about its underlying structure. In this approach, the phenomenon of consonant clustering therefore remains unintelligible.

The representation adopted in the Syllabified Consonant Analysis (Bush 1997) is empirically adequate, but on the theoretical level offers neither generality nor abstractness, which is desirable for any type of analysis. The lack of generality and abstractness is echoed in complex segmental structure, such as *prtx͡i* in (51b). This account fails on grounds of explanatory adequacy too, since one cannot see the reason for /prtxili/ having this unusually long consonant sequence. The strategy of the SCA is simply to consider the unusual as usual, without making any effort to understand the phenomenon of consonant clustering. The SCA, similarly to the HSA, is descriptive rather than explanatory.

The Government Phonology representation (Toft 1999) in (51c) fails on the basis of empirical adequacy. There is no phonological or phonetic motivation for positing three occurrences of schwa in /prtxili/. Other empirical considerations are not taken into account either (see section 6.3.1). Thus, the representation offered by Toft is both too abstract and too simple to start with. The oversimplification in the representation in (51c) poses questions as to how one can justify and motivate the relation between the representation *pərətəxili* and the word /prtxili/. This analysis offers no insight into the understanding of the phenomenon of long consonant sequences, neither in terms of their constituency nor in terms of their occurrence. Thus, like the previous analyses discussed above, this attempt is not explanatorily adequate.

The Gradual Consonant Analysis representation in (51d) is empirically adequate in that it incorporates all segments of the word /prtxili/, and, in addition, takes into consideration the optional appearance of the sonorant /r/ and the complex nature of the harmonic clusters. There is one extra element, the vowel represented as V, which makes the representation *p(r)Vt͡x-il.i* simple and abstract at the same time.

This representation is simple for two reasons: (i) the optional appearance of the sonorant /r/ is taken into account and (ii) the harmonic cluster *tx* is presented as a complex segment. In addition, the representation, unlike for instance Toft's representation (with empty constituents), does not include any extra structural constituents which are not motivated by language facts.

The representation is abstract because it includes the unspecified vowel, V, which is introduced based on comparative evidence (for the detailed account of the word in the GCA see Chapter 6).

The phonological string of the word /prtxili/ is interpreted in terms of morphological constituents, e.g. the stem and the affix, and the theory of phonotactics advocated in this thesis provides phonological principles and constraints which are operative on morphological domains. Thus, the well-formedness of consonant se-

quences is established and explained with reference to morphology (see Chapter 3). Finally, after taking into consideration two empirically justified conditions, one can see the relation between the representation *p(r)Vt͡x-il.i* and the word /prtxili/. These two conditions are: (i) the sonorant /r/ between consonants sharing a laryngeal specification is optional and (ii) a stem vowel deletes when followed by a vowel-initial suffix. If both conditions are met, this yields the word /prtxili/. Thus, the GCA is explanatory and has predictive power too. It operates with general terms such as the stem and the affix which are common linguistic theoretical constructs, combined with phonological principles, e.g. the SSP and the OCP, which are cross-linguistically well acknowledged too. Thus, the GCA provides a deeper understanding of the phenomenon of consonant sequences in Georgian.

The difference between the previous analyses and the GCA lies in the concepts they operate with. In the HSA, the SCA and Toft's analysis, the formal side of a language, phonotactics, is interpreted in terms of formal units: the syllable, the foot, the onset and the nucleus, while in the GCA, phonotactics is viewed in three dimensions: meaning, form and structure. Such an approach to the study of phonotactics bridges the gap between phonology and other modules of a language. This yields a better understanding of the phenomenon that has previously considered to be merely phonological. Hence, as shown above, the study of phonotactics gains explanatory power when morphology, for instance, is taken into account, and leaves formal structures unintelligible when it is ignored.

Another difference between the previous analyses and the GCA lies in the novel approach to phonotactics proposed in the GCA. Unlike previous studies, which regarded phonotactics as a static structural component of a language, the GCA, by using a different methodology, the comparative, historical and consideration of paradigmatically related forms, regards phonotactics as a dynamic component of language organisation.

7.4.2. Demonstration

The process by way of which the different types of analyses will be demonstrated is metathesis (see also Butskhrikidze & van de Weijer 2001a). Metathesis in Modern Georgian moves the sonorant /v/, which is part of the thematic suffix, into the root in verbal forms before the infinitival suffix /-a/. For example, the third person singular of the root /xar/ 'to gnaw' is /xravs/, from underlying /xar-av-s/ (with deletion of the root vowel). The infinitival form, however, is /xvra/, from underlying /xar-av-a/ (with deletion of both the root vowel and the vowel in the thematic suffix), where the thematic suffix consonant /v/ ends up between the two root consonants. More data are presented in (52).

(52) *Regular metathesis in verb forms*

root	present tense, 3SG (*-av* thematic suffix)	infinitival form (*-a* infinitival suffix)	
xar	xr-av-s /xar/-/av/-/s/	xvr-a /xar/-/av/-/a/	‘to gnaw’
k’ar	k’r-av-s	k’vr-a	‘to tie’
xan	xn-av-s	xvn-a	‘to plough’
k’al	k’l-av-s	k’vl-a	‘to kill’
sxal	sxl-av-s	sxvl-a	‘to chop off’
jer	jr-av-s	jvr-a	‘to move’

There are a few conditions on v-metathesis: the root has to end in a sonorant (so that metathesis does not occur with a root like /xed/ ‘to see’), and the root should not start with a labial (so that metathesis does not occur with a root like /ber/ ‘to blow up’). The former is related to the well-formedness of the onsets in Georgian, which prefer consonant + sonorant combinations. The latter condition is related to a constraint which bans two bilabials monomorphemically (see also Chapter 3). The metathesis facts in (52) illustrate that it is closely related to the vowel-reduction process. This is a widely attested process in verbal as well as nominal forms in Modern Georgian. There are several conditions in order for vowel reduction to occur in verbal roots.

(53) *Conditions on vowel deletion in verb roots*

a) The root contains one of the vowels /a e o/.
b) The root ends in a sonorant (explains (54a)).
c) The suffix has the shape -VC.
d) The well-formed cluster is formed.

The forms in (54) illustrate the conditions on metathesis as well as on vowel reduction. If vowel deletion does not occur, metathesis does not occur either.

(54) *Regular non-occurrence of vowel deletion and metathesis*

a)	xed	xed-av-s	xedv-a (*xvd-a)	‘to see’
	tes	tes-av-s	tesv-a (*tvs-a)	‘to sow’
	les	les-av-s	lesv-a (*lvs-a)	‘to sharpen’
b)	ber	ber-av-s	berv-a (*bvr-a)	‘to blow up’
	par	par-av-s	da-parv-a (*da-pvr-a)	‘to hide’
c)	lar	lar-av-s	lar-va (*lvr-a)	‘string’

In (54a), vowel reduction does not take place because the root does not end in a sonorant (see the condition in (53b)), and consequently metathesis is blocked. In (54b), vowel reduction does not take place because it would yield two bilabials in adjacency, so metathesis does not apply either. In (54c), the reduction would result in an ill-formed onset (see (53d)), and hence both processes remain inactive.

The question is whether any of the analyses discussed in this chapter can account for the metathesis facts of Georgian. Each analysis will be considered in turn.

The Headless Syllable Analysis (Nepveu 1994) employs some notion of the well-formedness of Georgian consonant sequences, so that it can account for the metathesis process exemplified in (52) by saying that according to constraints of Georgian phonotactics the sequence /k'^{w}l/ is better than /k'lv/. This would be the only reason for the metathesis process to occur in the HSA. However, this is merely part of the analysis of the metathesis process in Georgian. As shown above, the metathesis process is related to the phonotactic well-formedness of the stem (which cannot have two adjacent labials, for instance) and the reduction process, which is not considered in the HSA. Hence the HSA cannot explain why metathesis does not happen in the examples in (54). The HSA does not have the predictive power to ensure the occurrence of the process and thus cannot provide us with a comprehensive account of the metathesis facts.

Both the Syllabified Consonant Analysis (Bush 1997) and the analysis in Government Phonology (Toft 1999) fail to account for metathesis, because neither of them makes any predictions about the well-formedness of consonant sequences in Georgian. Both approaches take an extreme position by either stipulating that Georgian allows any kind of consonant sequence (Bush 1997) or that Georgian does not allow any branching onsets (Toft 1999). As shown in Chapter 3, Georgian allows maximally two-member consonant sequence word-initially, and this information is crucial for accounting for the metathesis process as described above (see (53)). In addition, neither analysis has anything to say about the well-formedness of morphological constituents or about the vowel-reduction process, which are both essential for understanding the metathesis facts. Hence, the process remains unaccounted for within both approaches.

In addition, within OT, Nepveu (1994) and Bush (1997) could invoke morphological information by introducing alignment constraints and other constraints sensitive to morphology. In fact, Nepveu uses alignment constraints to account for the distribution of headless syllables, but alignment constraints alone are not enough. Additional constraints on the phonotactic well-formedness of morphological constituents are also needed. Such constraints would ensure the occurrence of vowel deletion (see (53)), which is a necessary condition for the metathesis process to occur.

To conclude, in order to account for the metathesis process it is important to take into account several aspects: the well-formedness of consonant sequences in Georgian in terms of their constituency and distribution; the well-formedness of the phonotactics of stem and affix, a mechanism for the reduction process and behaviour of specific phonemes (in this case the behaviour of the sonorant /v/). The GCA considers all of these aspects and their interactions, and can explain and predict the occurrence of metathesis in Georgian (for more about analysis of metathesis in Georgian see Butskhrikidze & van de Weijer 2001a).

8 Summary and Conclusions

The study of consonant sequences, e.g. forms of the CCC type, is the central topic of this thesis. Assuming that language is an open system and anticipating the idea of relativity and complementariness, the hypothesis that CCC and CVCVCV are related is proposed. More specifically, I argue that all well-formed consonant sequences are derived from structures of the CVC type, which is the lexical part of a word, the stem. On the representational side, to account for the consonant phonotactics a phonological hierarchy is introduced in which the stem domain occupies a place between the segment and the word domain. By encompassing formal, meaning and structural components of a language, a three-dimensional approach is proposed to account for the word-level phonotactics. The linear sequences of segments of a word constitute its formal aspect, morphological constituency comprises the meaning aspect of a word and phonological principles (e.g. the Sonority Sequencing Principle, the Obligatory Contour Principle, among others, which are referred to as different instantiations of a single phenomenon, namely the Balancing Principle) constitute the structural aspect of the word. The main body of the thesis is devoted to testing the predictions and the hypotheses proposed in Chapters 1 and 2.

Due to their length and constituency, the consonant sequences of Georgian (e.g. /prckvn/, /mc'vrtn/, /brt'χ'/ in words such as /prckvna/ 'to peel', /mc'vrtneli/ 'trainer', /brt'χ'eli/ 'flat') appear to be problematic for researchers who assume that language is a system. The phonotactics and behaviour of the consonants in isolation, and the phonotactics and peculiarities of consonant clusters, are studied from synchronic, historical and comparative perspectives. In Chapter 3, it is established that there are three types of consonant sequences in Georgian which could be treated as complex segments: (i) harmonic clusters, (ii) sequences of C + /v/ and (iii) sequences of /s/ + obstruent. The minimal word in Georgian is disyllabic and consonant sequences are restricted to the stem domain. The hypothesis concerning the derived nature of consonant sequences is tested by comparing the patterns of consonant combinations in adjacency and across a vowel in a stem. A striking resemblance is found between the two, confirming the prediction that maximally two-member consonant sequences are derived from the stem domain. They are the result of stem-vowel deletion, a common process in both Old and Modern Georgian. It is demonstrated that consonantal complexity in Modern Georgian is due to morphological complexity, as well as the result of processes of vowel deletion and complex segment formation.

In Chapter 4 additional pieces of evidence are explored to support the claims made in Chapter 3. A perceptual technique is used to test the phonological claim

about treating harmonic clusters as complex segments. The experiment was conducted with 24 native Georgian listeners. Two dependent variables were used to evaluate the experimental results, viz. detection rate (hit rate in %) and detection latency (reaction time in ms). The experiment substantiates the phonological claim that harmonic clusters are complex segments.

Reduplication generally reveals the unmarked structures of a language, and the reduplication data of Georgian presented in Chapter 5 support the claims made in Chapter 3: (i) The minimal word in Georgian is disyllabic; (ii) harmonic clusters can be analysed as complex segments; (iii) C + /v/ combinations can be analysed as complex segments; (iv) /s/ + obstruent can be analysed as complex segment; (v) obstruent + sonorant is the most unmarked consonant cluster; and (vi) Georgian maximally allows two-member consonant clusters stem-initially. Consequently, the data support the claim that consonant sequences are, in general, derived.

The last two chapters concern analyses of Georgian consonant sequences. The Gradual Consonant Analysis (GCA) is proposed in Chapter 6. The analysis incorporates the following types of evidence: paradigmatic, syntagmatic, phonetic, historical and comparative. Examination of consonantal stems of Georgian and the application of the GCA shows that language-external and internal evidence provide a sound basis for exploring the nature of 'complex' structures of a language. In most cases, the long consonant sequences appear to be the result of a very productive vowel-reduction process and of complex segment formation. Georgian is an inflectional-agglutinative language and consonant sequences are generally derived from CVC stems when vowel-initial affixes are added. Thus, the phonological 'complexity' is primarily due to morphological complexity and the structures of the CCC type in Georgian appear to be the result of transformations of the structures of the CVCVCV type, as suggested in Chapters 1 and 2.

In Chapter 7, previous analyses of Georgian consonant sequences are reviewed: the Headless Syllable Analysis (HSA), the Syllabified Consonant Analysis (SCA) and an analysis within the framework of Government Phonology (Toft 1999). Several criteria are used to compare the GCA to the previous analyses: (i) simplicity or economy, (ii) generality or abstractness and (iii) adequacy (empirical and explanatory). The evaluation of previous analyses and the GCA against these criteria shows that the GCA provides a better understanding of Georgian consonant sequences.

The differences between the previous analyses and the GCA lie in the concepts they operate with. In the HSA, the SCA and in Toft's (1999) analysis, the formal side of a language, phonotactics, is interpreted in terms of formal units: the syllable, the foot, the onset and the nucleus. In the GCA, however, phonotactics is viewed in three dimensions: meaning, form and structure. Such a novel approach to the study of phonotactics links phonology with other modules of language. This yields a better understanding of phenomena that were before considered merely phonological. The phonotactic study gains explanatory power when morphology is taken into account.

Another difference between the previous analyses and the GCA lies in a novel approach to phonotactics proposed in the GCA. Unlike previous studies, which re-

garded phonotactics as a static structural component of a language, the GCA, by using a different methodology (including comparative and historical) regards phonotactics as a dynamic component of language organisation.

Thus, the following claims are made in this thesis:

- Phonotactics is dynamic.
- Phonotactics is sensitive to morphology.
- The stem is a phonological domain.
- Consonantal complexity in Modern Georgian is due to morphological complexity, as well as the result of processes of vowel deletion and of complex segment formation.

Appendix 1:
Georgian Nominal and Verbal Forms

In this appendix, the lists of Modern Georgian nominal and verbal 'stem types' given by Ertelishvili (1970, 1980) are considered. Ertelishvili points out that the most common Georgian stem type is monosyllabic. Disyllabic and trisyllabic stems are usually loan words or morphologically derived, e.g. compounds, reduplicated forms, etc. It is important to note that, according to Ertelishvili, the stem can be morphologically complex. In this book, the stem refers to a single lexical morpheme, not to derived forms. Thus, these lists include both monomorphemic and polymorphemic nominal and verbal forms of Georgian.

1. Nominal stems (Ertelishvili 1980)

a) 2764 monosyllabic nominal stems are analysed; 25 patterns are distinguished.

Number	Nominal Stems	Number of occurrences	Percentage
1	CVCC	760	27.50
2	CVC	715	25.87
3	CCVC	437	15.81
4	CCVCC	197	7.13
5	CVCCC	141	5.10
6	CCCCVC	127	4.59
7	CCV	75	2.71
8	VCC	50	1.81
9	CCCV	50	1.81
10	CCVCCC	38	1.37
11	CCCCVC	31	1.12
12	VC	30	1.08
13	CCCVCC	30	1.08
14	CVCCCC	24	0.87
15	CV	20	0.72
16	CCCCV	9	0.33
17	VCCC	8	0.29
18	CCCVCCC	5	0.18

19	VCCCC	4	0.14
20	CCVCCCC	1	0.04
21	CCCCCVC	1	0.04
22	CVCCCCC	1	0.04
23	CCCCVCC	1	0.04
24	CCCVCCCC	1	0.04
25	CCCCVCCC	1	0.04

b) 5166 disyllabic nominal stems are analysed; 54 patterns are distinguished.

Number	Nominal stems	Number of occurrences	Percentage
1	CVCVC	1380	26.71
2	CVCCVC	1020	19.74
3	CVCV	610	11.81
4	CVCCV	480	9.29
5	CCVCV	179	3.46
6	CCVCVC	169	3.25
7	VCCVC	161	3.12
8	CVCVCC	149	2.88
9	CVVC	130	2.52
10	CVCCCVC	125	2.42
11	CVVVCVV	110	2.13
12	CVCCCV	78	1.51
13	CVCCCV	62	1.20
14	CCVCCV	61	1.18
15	CCVCCVC	60	1.16
16	CVV	42	0.81
17	VCCV	35	0.68
18	VCV	30	0.58
19	CCVVC	30	0.58
20	VCVCC	26	0.50
21	VVC	22	0.43
22	VCCVCC	22	0.43
23	CCVCVCC	21	0.41
24	VCCCVC	20	0.39
25	CCCVCV	18	0.35
26	CCVV	17	0.33
27	CCCVCVC	14	0.27
28	CVVCC	12	0.23
29	CVCCCCVC	10	0.19
30	VVCC	9	0.17
31	CCVCCVCC	8	0.15
32	CVCCCVCC	8	0.15

33	VCCCVCC	6	0.11
34	CCVCCCVC	5	0.10
35	CVCVCCC	4	0.08
36	CCCVV	3	0.06
37	CCVVCC	3	0.06
38	CCCCVCV	3	0.06
39	CCCVCCV	3	0.06
40	CCVCCCVC	3	0.06
41	VCCCV	2	0.04
42	VCCCCV	2	0.04
43	CCCVVC	2	0.04
44	CCCVCCVC	2	0.04
45	CCVCCCCV	2	0.04
46	VV	1	0.02
47	CCCCVV	1	0.02
48	CCCCVC	1	0.02
49	CCCVVCC	1	0.02
50	VCCCCVC	1	0.02
51	VCCVCCC	1	0.02
52	VCCCCVCC	1	0.02
53	CCVCCCVCC	1	0.02
54	CCVCCCCCVC	1	0.02

2. Structural models of verbal stems (Ertelishvili 1970)

Number	Verbal stems	Number of occurrences	Percentage

a) Non-syllabic stems

1	CC	45	0.821
2	CCC	42	0.766
3	CCCC	14	0.255
4	C	14	0.255

b) Monosyllabic stems

5	CCVC	815	14.870
6	CVC	725	13.227
7	CVCC	620	11.312
8	CCCVC	300	5.473
9	CVCCC	250	4.561
10	CCVCC	240	4.379
11	CCCCVC	80	1.460
12	CCCVCC	50	0.912

13	CCVCCC	42	0.766
14	CVCCCC	36	0.657
15	CCCCCVC	17	0.310
16	CV	14	0.255
17	CCV	11	0.201
18	VC	10	0.182
19	VCC	9	0.164
20	Other types	5	0.091

c) Disyllabic stems

21	CVCCVC	710	12.954
22	CVCVC	400	7.298
23	CCVCVC	350	6.386
24	CVCCCVC	105	1.919
25	CCVCCVC	70	1.277
26	CCVCCCVC	55	1.003
27	CCCVCVC	45	0.821
28	CCCVVC	45	0.821
29	CVVC	40	0.730
30	CCVVC	40	0.730
31	CCCCVVC	18	0.328
32	CCCVCCVC	10	0.182
33	CCCCCVVC	7	0.128

d) Trisyllabic stems

34	CCVCV-CCVC	40	0.730
35	CVCV-CVC	40	0.730
36	CVCVVC	40	0.730
37	CVCCVCVC	30	0.547
38	CVCVCVC	28	0.511
39	CVCCV-CVCC	17	0.310
40	CCVCVVC	13	0.237
41	CVCCVVC	12	0.219
42	CCCVCV-CCCVC	8	0.146
43	Different types	19	0.347

Appendix 2:
Stem-initial Consonant Sequences

This is a list of consonant sequences (two-, three-, four-, five- and six-member sequences) that appear in stem/word-initial position in Georgian. The data come from Žgenti (1956). I have changed some of the examples and have added a number of sequences that are very common in Georgian. The list illustrates patterns of consonant sequences and is not exhaustive. It does not contain sequences of the /m/ + C type, and there are only two cases of sequences of the /v/ + C type. The first column lists well-formed consonant combinations, the second lists Georgian words in which the combinations are attested and the third lists glosses.

1. Two-member (CC) consonant sequences

a) stop + stop

bg	bgera	‘sound’
dg	dguši	‘sucker’
gd	gdeba	‘lie about’
k’b	k’bili	‘tooth’
p’k’	p’k’ureba	‘spray’
tb	tbili	‘warm’
tk	tkeši	‘downpour’
t’k’	t’k’acani	‘crack’

b) stop + affricate

brǰ	brǰeni	‘rest’
bč’	bč’oba	‘discussion’
brj	brjaneba	‘order’
kc	kceva	‘behaviour’
t’χ’	t’χ’avi	‘skin’

c) stop + fricative

bz	bza	‘box-tree’
bž	bžit’i	‘duckling’
bɣ	bɣavili	‘bleating’
dɣ	dɣe	‘day’
gz	gza	‘road’

	ks	kseli	‘cob-web’
	kš	kšena	‘to puff’
	pš	pšat’i	‘type of a plant’
	px	pxa	‘skill’
	tx	txa	‘goat’
d) stop + sonorant			
	bl	blu	‘tongue-tied’
	bm	bma	‘to bind’
	bn	bneva	‘fastening’
	br	brazi	‘evil’
	dv	dvire	‘log’
	dm	dmanisi	‘toponym’
	dn	dnoba	‘melting’
	dr	dro	‘time’
	gv	gvalva	‘drought’
	gl	glova	‘grief’
	gm	gmiri	‘hero’
	gn	gnomi	‘gnome’
	gr	gradus i	‘degree’
	kv	kva	‘stone’
	kl	kliavi	‘plum’
	km	kmediti	‘effective’
	kn	knari	‘lyre’
	kr	kris	‘to whiff’
	k’v	k’veba	‘to feed’
	k’l	k’lanč’i	‘claw’
	k’m	k’maχ’opili	‘content’
	k’n	k’nut’i	‘kitten’
	k’r	k’reba	‘meeting’
	pl	ploba	‘possess’
	pr	prinveli	‘bird’
	p’l	p’lombi	‘stopping’
	p’r	p’riali	‘sparkling’
	tv	tvali	‘eye’
	tl	tla	‘peel’
	tm	tma	‘hear’
	tr	treva	‘drag’
	t’v	t’vini	‘brain’
	t’l	t’lanki	‘rough’
	t’r	t’rabaxi	‘boasting’
e) affricate + stop			
	cd	cduneba	‘temptation’
	ck	ckera	‘look’

čk	čkari	‘fast’
c’k’	c’k’aruni	‘tickling’
č’d	č’de	‘notch’
ǰd	ǰdoma	‘to sit’
ǰg	ǰgupi	‘group’

f) affricate + fricative

cx	cxadi	‘obvious’
čx	čxik’vi	‘jay’
c’χ’	c’χ’ali	‘water’
č’χ’	č’χ’int’i	‘new (cheese)’
jɣ	jɣola	‘accompany’
ǰɣ	ǰɣani	‘patch’

g) affricate + sonorant

cv	cvari	‘dew’
cm	cmacuni	‘to move lips’
cn	cneba	‘notion’
cr	cremli	‘tear’
čv	čvili	‘infant’
čl	čliki	‘hoof’
čr	čra	‘cut’
c’v	c’va	‘burning’
c’l	c’liuri	‘annual’
c’m	c’mida	‘holy’
c’n	c’nexa	‘pressing’
c’r	c’re	‘circle’
č’v	č’vavi	‘rye’
č’l	č’leki	‘tuberculosis’
č’m	č’meva	‘feeding’
č’r	č’reli	‘motley’
jv	jvali	‘bone’
jl	jlieri	‘strong’
jm	jma	‘brother’
jn	jna	‘sheaf’
jr	jroxa	‘cow’
ǰv	ǰvari	‘cross’
ǰm	ǰmuxi	‘stumpy’

h) fricative + stop

sk	skesi	‘gender’
sk’	sk’a	‘hive’
st’	st’umari	‘guest’
šp	špoti	‘trouble’
št	štantkma	‘absorption’
št’	št’o	‘branch’

xd	xdoma	‘to lose flesh’
χ’b	χ’ba	‘jaw’
χ’d	χ’da	‘book cover’

i) fricative + affricate

sc’	sc’avla	‘study’
xc	xceneba	‘mention’

j) fricative + fricative

sx	sxivi	‘beam’
šx	šxap’una	‘pouring (rain)’
zɣ	zɣap’ari	‘tale’
žɣ	žɣera	‘phonation’
xš	xširi	‘frequent’

k) fricative + sonorant

sv	svavi	‘griffon’
sl	slok’ini	‘hiccup’
sm	smena	‘hearing’
sr	srola	‘shooting’
šv	švili	‘child’
šl	šlami	‘silt’
šm	šmori	‘mustiness’
šn	šno	‘charm’
šr	šriali	‘rustle’
zv	zvavi	‘avalanche’
zl	zlazvna	‘sluggishness’
zm	zmuili	‘lowing’
zn	zne	‘character’
zr	zrunva	‘care’
žr	žriamuli	‘hub’
žl	žlet’a	‘to exterminate’
ɣv	ɣvino	‘wine’
ɣl	ɣlet’a	‘fleece’
ɣm	ɣmeč’a	‘to make a face’
ɣr	ɣreoba	‘carouse’
xv	xvadi	‘male’
xl	xleba	‘accompany’
xm	xma	‘voice’
xn	xnieri	‘elderly’
xr	xrinc’i	‘wheeze’
χ’v	χ’vavili	‘flower’
χ’l	χ’lort’i	‘sprout’
χ’m	χ’muili	‘howling’
χ’n	χ’nosva	‘to smell’
χ’r	χ’ra	‘to spread over, throw’

l) sonorant + stop

lb	(da)lboba	‘to soak’
lt’	lt’olva	‘aspiration’
nd	ndoba	‘faith’
rb	rbeva	‘to raid’
rg	rgoli	‘ring’
rk	rka	‘horn’
rk’	rkina	‘iron’
rt’	rt’o	‘branch’

m) sonorant + affricate

rc’	rc’eva	‘rocking’
rj	rjali	‘daughter-in-law’
rǰ	rǰuli	‘religion’

n) sonorant + fricative

vs	vseba	‘to fill’
lx	lxena	‘merriment’
rχ’	rχ’eva	‘fluctuation’

o) sonorant + sonorant

lm	lmobieri	‘soft-hearted’

2. Three-member (CCC) consonant sequences

a) stop + stop + stop

t’k’b	t’k’bili	‘sweet’

b) stop + stop + sonorant

tkv	tkveni	‘your’
tkr	tkriali	‘gush out’

c) stop + affricate + stop

p’c’k’	p’c’k’ari	‘line’

d) stop + fricative + stop

psk’	psk’eri	‘bottom’

e) stop + fricative + sonorant

bɣv	bɣvera	‘frown’
txl	txle	‘sediment’
txr	txroba	‘to tell’
t’χ’v	t’χ’via	‘bullet’
t’χ’l	t’χ’lap’i	‘fruit cookie’

f) stop + sonorant + stop

brb	brbo	‘crowd’
brg	brke	‘mould’
brk	brge	‘stalwart’
grd	grdemli	‘anvil’
krt	krtami	‘bribe’
k’vd	k’vdoma	‘to die’
k’ld	k’lde	‘rock’
k’rt	k’rtoma	‘trembling’
prt	prta	‘wing’
trt	trtolva	‘trembling’
zrd	zrda	‘growth’

g) stop + sonorant + affricate

grj	grjeli	‘long’
k’rj	k’rjalva	‘reverence’

h) stop + sonorant + sonorant

brm	brma	‘blind’
gvr	gvrit’i	‘turtle-dove’
k’vn	k’vniti	‘piece’
k’vr	k’vra	‘to push’
tvl	tvla	‘count’

i) affricate + stop + sonorant

ckv	ckvit’i	‘frisky’
c’k’l	c’k’lap’uni	‘champing’
c’k’m	c’k’muili	‘whine’
c’k’r	c’k’riali	‘tickle’
č’k’n	č’k’noba	‘fade’

j) affricate + fricative + sonorant

cxv	cxviri	‘nose’
cxr	cxra	‘nine’
čxr	čxrek’a	‘search’
c’χ’v	c’χ’vili	‘couple’
c’χ’l	c’χ’luli	‘sore’
c’χ’r	c’χ’roma	‘anger’
c’χ’n	c’χ’nari	‘quiet’
č’χ’l	č’χ’let’a	‘crush’
jɣv	jɣveni	‘present’

k) affricate + sonorant + stop

c’rt	c’rtoba	‘hardening’
c’rp	c’rpeli	‘sincere’

l) affricate + sonorant + affricate

jrc’	jrc’ola	‘tremble’

m) affricate + sonorant + fricative

črd	črdili	‘shade’

n) affricate + sonorant+ sonorant

c’vr	c’vrili	‘small’
č’vr	č’vret’a	‘contemplation’
jvr	jvra	‘move’

o) fricative + stop + sonorant

st’v	st’vena	‘whistle’

p) fricative + affricate + sonorant

sc’r	sc’rapi	‘quick’

q) fricative + fricative + sonorant

sxv	sxva	‘other’
zɣv	zɣva	‘sea’
xsn	xsnari	‘solution’

r) fricative + sonorant + stop

xrt’	xrt’ili	‘gristle’
χ’rd	χ’rdena	‘to lean’

s) fricative + sonorant + affricate

ɣrj	ɣrjili	‘gum’

t) fricative + sonorant + fricative

žrž	žržola	‘shivering’

u) fricative + sonorant + sonorant

švr	švria	‘oats’
zmn	zmna	‘verb’
ɣrm	ɣrma	‘deep’
xvr	xvret’a	‘to make a hole’

v) sonorant + stop + sonorant

rgv	rgva	‘to plant’
rtv	rtveli	‘vintage’

w) sonorant + affricate + sonorant

njr	njreva	‘reeling’

x) sonorant + affricate + fricative

rcx	rcxila	‘hornbeam’
rc’χ’	rc’χ’eva	‘vomiting’

y) sonorant + fricative + sonorant

rɣv	rɣveva	‘disorganisation’

z) sonorant + sonorant + affricate

vrc	vrceli	‘vast’

3. Four-member (CCCC) consonant sequences

a) stop + stop + sonorant + sonorant

tkvl	tkvlepa	‘lap’

b) stop + fricative + fricative + sonorant

txzv	txzva	‘compose’

c) stop + fricative + sonorant + sonorant

pxvn	pxvnili	‘powder’
pšvn	pšvnet’a	‘to rub one’s hands’

d) stop + sonorant + stop + fricative

brt’χ’	brt’χ’eli	‘flat’
prtx	prtxili	‘careful’

e) stop + sonorant + stop + sonorant

grgv	grgvinva	‘thunder’
prkv	prkveva	‘to shed’

f) stop + sonorant + affricate + fricative

brc’χ’	brc’χ’invaleba	‘brilliance’
brč’χ’	brč’χ’ali	‘claw’
prčx	prčxili	‘nail’

g) stop + sonorant + affricate + sonorant

grjn	grjnoba	‘feeling’

h) affricate + fricative + sonorant + stop

c’χ’vd	c’χ’vdiadi	‘darkness’

i) affricate + fricative + sonorant + sonorant

čxvl	čxvlet’a	‘prick’

j) affricate + sonorat + stop + sonorant

c’rtv	c’rtvna	‘training’

k) fricative + fricative + sonorant + sonorant

sxvl	sxvla	‘chop off’

l) fricative + sonorant + affricate + sonorant

xrc’n	xrc’na	‘decay’

m) fricative + sonorant + fricative + sonorant

ɣrɣn	ɣrɣna	‘grumbling’

n) sonorant + stop + fricative + sonorant

rt’χ’m	rt’χ’ma	‘to hit’

o) sonorant + affricate + fricative + sonorant

nǰɣr	nǰɣreva	‘to shake’
rcxv	rcxvena	‘shame’

p) sonorant + fricative + sonorant + sonorant

rχ’vn	rχ’vna	‘corrupt’

4. Five-member (CCCCC) consonant sequences

a) affricate + sonorant + sonorant + stop + sonorant

c’vrtn	c’vrtna	‘training’

5. Six-member (CCCCCC) consonant sequences

a) stop + sonorant + affricate + stop + sonorant + sonorant

prckvn	prckvn	‘to peel’
brdɣvn	brdɣvna	‘to fight’

Appendix 3:
Stem-final Consonant Sequences

This is a list of consonant sequences (two-, three-, four- and five-member sequences) that appear in stem-final (i.e. word-medial) position in Georgian. The list illustrates patterns of consonant sequences and is not exhaustive.

1. Two-member (CC) consonant sequences

The data on two-member stem-final consonant sequences are from Uturgaidze (1976). The sequences commonly found in stem-final position are either harmonic clusters or the sequences of the obstruent + sonorant type. The same sequences are found in stem-initial position too, e.g.

a) Harmonic clusters
{bɣ dɣ zɣ žɣ px sx šx c'χ' č'χ'}

b) Obstruent + sonorant
{zv zr zl pr tr sl šl žr šn zn xl xr ɣl ɣr}

Clusters which are attested in the stem-initial position but not stem-finally are:

{t'b k'b xb tb gd xd gj}

Interestingly, these are non-deccessive clusters, which are also not characteristic of stem-initial position. They are found only in a few words. In Chapter 3, it is shown that most of these clusters are secondary.

Clusters which are attested neither in stem-initial nor in stem-final position are:

{zb xp χ'p' žb rp rž lž lš lj}

Sequences of the sonorant + obstruent type are very common, but, interestingly, most of them are loan words. The matter is beyond the scope of this thesis, but certainly deserves further investigation.

nd	gund-i (Persian)	‘choir’
nj	sp’ilenj-i (Persian)	‘copper’
nǰ	brinǰ-i (Persian)	‘rice’
ng	niangi (Persian)	‘crocodile’
nk’	mank’-i (Persian)	‘defect’
nč	kanč-i (Turkish)	‘(screw-) nut’
rg	barg-i (Persian)	‘luggage’
rd	vard-i (Persian)	‘rose’
rǰ	xarǰ-i (Arabic)	‘expense’
rč	čarč-i (Turkish)	‘profiteer’
rk’	xark’-i (Arabic)	‘tribute’
rp’	k’erp’-i (Persian)	‘idol’
rj	k’erj-i (Persian)	‘meal’
rp	šarp-i (French)	‘scarf’
rb	xarb-i (Persian)	‘greedy’
lk	olk-i (Turkish)	‘district’
lx	xalx-i (Arabic)	‘people’
vz	k’ovz-i (Persian)	‘spoon’
vt	navt-i (Persian)	‘kerosene’

Obstruent sequences attested only in loan words are:

{st’ št’ zd xt’ šk’ xš ɣd}

2. Three-member (CCC) consonant sequences

Most of the three-member sequences are of the three types: (i) harmonic cluster + sonorant, (ii) sonorant + harmonic cluster and (iii) sonorant + obstruent + sonorant.

a) harmonic cluster + sonorant

pxv	vepxv-i	‘tiger’
cxl	cecxl-i	‘fire’
cxv	ricxv-i	‘number’
t’χ’l	mat’χ’l-i	‘wool’
sxl	sisxl-i	‘blood’

b) sonorant + harmonic cluster

nčx	čončx-i	‘skeleton’
rcx	marcx-i	‘failure’
	parcx-i	‘harrow’
rt’χ’	bart’χ’-i	‘nestling’
všv	bavšv-i	‘child’

c) sonorant + obstruent + sonorant

rpl	parpl-i	‘fin’
	perpl-i	‘ashes’
rp’l	msxerp’l-i	‘victim’
rbl	zɣurbl-i	‘threshold’
rxv	verxv-i	‘asp’
rɣv	jarɣv-i	‘vein’
rcv	borcv-i	‘hill’
rt’l	zɣmart’l-i	‘medlar’
rjl	ɣvarjl-i	‘spite’
lkv	bolkv-i	‘bulb’
nc’l	anc’l-i	‘elder’

3. Four-member (CCCC) consonant sequences

Four-member sequences usually contain harmonic clusters and two sonorants. Similarly to the stem-initial clusters, the first sonorant of the sequences is usually /r/, which is followed by a harmonic cluster and the sonorant /v/.

rtkl	ortkl-i	‘steam’
rcxl	vercxl-i	‘silver’
rcxv	morcxv-i	‘shy’
rc’χ’v	marc’χ’v-i	‘strawberry’
nčxl	ančxl-i	‘irritable’

4. Five-member (CCCCC) consonant sequences

Five-member sequences are very rare and contain harmonic clusters and three sonorants.

nǰgvl	banǰgvl-i	‘shaggy’

Appendix 4:
Experimental Data

This is a list of the full set of targets and stimuli (non-words) used in the experiment described in Chapter 4.

k'ab	k'a.bga.na	tej	te.jgi.mo	bez	be.zgu.sa
	k'ab.ra.na		tej.di.mo		bez.ru.sa
	k'a.ba.na		te.ji.mo		be.zu.sa
	a.ba.na		ge.ri.mo		je.gu.sa
	ma.sa.na		se.di.mo		de.bu.sa
	ba.ra.na		le.ki.mo		le.mus.a
dap	da.pki.nu	c'aj	c'a.jgu.li	nas	na.sxa.zi
	dap.ri.nu		c'aj.nu.li		nas.ma.zi
	da.pi.nu		c'a.ju.li		na.sa.zi
	ta.ri.nu		da.su.li		ma.la.zi
	da.k'i.nu		da.c'u.li		k'a.pa.zi
	da.di.nu		da.zu.li		ja.ra.zi
sad	sa.dge.ri	maǰ	ma.ǰgu.di	zaž	za.žgu.lu
	sad.le.ri		maǰ.ru.di		zaž.ku.lu
	sa.de.ri		ma.ǰu.di		za.žu.lu
	va.de.ri		ga.ru.di		la.su.lu
	ga.me.ri		ja.χ'u.di		χ'a.du.lu
	ja.ne.ri		la.ku.di		ta.xu.lu
c'ad	c'a.dge.ri	maǰ	ma.ǰga.li	dap'	da.p'k'a.li
	c'ad.le.ro		maǰ.sa.li		dap'.ra.li
	c'a.de.ri		ma.ǰa.li		da.p'a.li
	va.ge.ri		na.ga.li		ba.ta.li
	ǰa.xe.ri		ka.la.li		k'a.da.li
	sa.te.ri		da.za.li		ra.da.li

cap	ca.pxi.li	lač	la.čxa.di	lat’	la.t’k’a.ni
	cap.ri.li		lač.ma.di		lat’.ra.ni
	ca.pi.li		la.ča.di		la.t’a.ni
	pa.ni.li		k’a.sa.di		k’a.ja.ni
	da.vi.li		pa.za.di		ga.da.ni
	na.di.li		na.p’a.di		sa.ca.ni
dat	da.tko.mi	t’as	t’a.sko.ni	zat’	za.t’χ’a.di
	da.tro.mi		t’as.mo.ni		zat’.ra.di
	da.to.mi		t’a.so.ni		za.t’a.di
	ga.so.mi		da.xo.ni		sa.ma.di
	k’a.do.mi		ba.go.ni		ja.sa.di
	na.no.mi		ta.lo.ni		k’a.sa.di
nat	na.txi.li	gež	ge.žgu.k’a	dec’	de.c’k’e.vi
	nat.mi.li		gež.du.k’a		dec’.de.vi
	na.ti.li		ge.žu.k’a		de.c’e.vi
	ša.di.ri		de.xu.k’a		ge.xe.vi
	k’a.ci.li		me.su.k’a		ne.ge.vi
	ma.ni.li		c’e.zu.k’a		te.ge.vi
k’ac	k’a.cki.li	χ’oš	χ’o.šxo.pi	vec’	ve.c’χ’u.k’i
	k’ac.gi.li		χ’oš.mo.pi		vec’.gu.k’i
	k’a.ci.li		χ’o.šo.pi		ve.c’u.k’i
	da.ci.li		ro.k’o.pi		ge.du.k’i
	sa.ri.li		do.no.pi		se.xu.ki
	ža.di.li		p’o.go.pi		ze.su.k’i
dac	da.cxo.mi	tab	ta.bga.mi	joč	jo.č’k’o.li
	dac.k’o.mi		tab.ra.mi		joč’.mo.li
	da.co.mi		ta.ba.mi		jo.č’o.li
	ga.do.mi		ča.ra.mi		do.ro.li
	k’a.so.mi		da.ra.mi		p’o.so.li
	la.do.mi		ta.ra.mi		do.do.li
seč	se.čko.na	čap’	ča.p’χ’a.ri	sač’	sa.č’χ’u.gi
	seč.lo.na		čap’.la.ri		sač’.ru.gi
	se.čo.na		ča.p’a.ri		sa.č’u.gi
	š e.bo.na		da.p’a.li		da.cu.gi
	ge.so.na		sa.ra.li		va.nu.gi
	p’e.lo.na		ša.ma.li		la.su.gi

gez	ge.zgo.bi	beš	be.ško.mi
	gez.m o.bi		beš.lo.mi
	ge.zo.bi		be.šo.mi
	se.do.bi		ge.so.mi
	xe.bo.bi		pe.do.mi
	le.do.bi		de.go.mi

References

Abdaljan, I.P. (1978). *Fonologičeskoe stroenije sluzhebnyx morfem*. [The phonological make-up of grammatical morphemes]. Doctoral dissertation, University of Moscow.

Aikhenvald, A.Y. (1996). Words, phrases, pauses and boundaries: evidence from South American Indian languages. *Studies in Language* **203**, 487–517.

Akhvlediani, G. (1949). *Zogadi ponet'ik'is sapudzvlebi*. [Foundations of general phonetics]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.

Allen, W.S. (1953). *Phonetics in ancient India*. *London oriental series* **1**. London: Oxford University Press.

Anderson, J.M. (1969). Syllabic or non-syllabic phonology. *Journal of Linguistics* **5**, 136–143.

Anderson, J.M. & C.J. Ewen (1987). *Principles of dependency phonology*. Cambridge: Cambridge University Press.

Archangeli, D. (1986). The OCP and Nyangamurda buffer vowels. *North Eastern Linguistic Society* **16**, 24–46.

Árnason, K. (1985). Morphology, phonology and u-umlaut in Modern Icelandic. Phono-morphology. *Studies in the interaction of phonology and morphology*, edited by E. Gussmann, 9–23. Lublin: Redakcja Wydawnictw.

Aronoff, M. (1994). *Morphology by itself*. Cambridge, Massachusetts: MIT Press.

Aronson, H.I. (1990). *Georgian*. *Reading grammar*. Columbus: Slavica Publishers, Inc.

Asatiani, R. (1983). Marcvlovani tankhmovnebi saerto-kartvelurshi. [The syllabic consonants in Proto-Kartvelian]. *Jevanmardi* **3**, 9–15.

Avrorin, V.A (1959). *Grammatica nanaijskogo jazyka*. [A grammar of Nanay]. Moscow-Leningrad: Izdatel'stvo Akademii Nauk SSSR.

Bade, M.V. (1986). *Sootnošenie morfologičeskogo i fonetičeskogo členenija reči i ejo vosprijatije*. [The correlation between the morphological and phonetic dismembering of speech and its perception]. Doctoral dissertation, University of Moscow.

Bagemihl, B. (1991). Bella Coola syllable structure. *Linguistic Inquiry* **22**, 589–646.

Bakradze, D. (1889). *Ist'oria sakartvelosi (udzvelesi droidam me-10 sauk'unis dasasrulamde).* [A history of Georgia], Vol. 1. Tbilisi: St'amba Meskhisa.

Bakró-Nagy, M.Sz. (1992). *Proto-phonotactics. Phonotactic investigation of the PU and PFU consonant system. Studia Uralica* **5**. Wiesbaden: Harrassowitz Verlag.

Baudouin de Courtenay, J. (1963). *Izbrannye trudy po obščemy jazykoznaniju.* [Selected writings on general linguistics]. Moscow: Akademija Nauk SSSR.

Bauman, M. (1995). *The production of syllables in connected speech*. Doctoral dissertation, Nijmegen University.

Bell, A. (1976). The distributional syllable. *Linguistic studies offered to Joseph Greenberg* **2**, 249–262. Saragota, California: Anma Libri.

Bell, A. (1978). Syllabic consonants. *Universals of human language* **2**, *Phonology*, edited by J.H. Greenberg, 153–193. Stanford: Stanford University Press.

Bell, A. & J. Bybee Hooper (1978). Issues and evidence in Syllabic Phonology. *Syllables and segments*, edited by A. Bell & J. Bybee Hooper, 3–25. Amsterdam: North-Holland Publishing Company.

Belova, A.G. (1991). Struktura semitskogo kornja i semitskaja morfologičeskaja sistema. [The Semitic root structure and the Semitic morphological system]. *Voprosy Jazykoznanija* **1**, 79–90.

Blevins, J. (1995). The syllable in phonological theory. *The handbook of phonological theory*, edited by J.A. Goldsmith, 206–244. Oxford: Blackwell.

Bloomfield. L. (1933). *Language*. New York: Holt.

Boersma, P. & D. Weenink (1996). Praat, a System for doing Phonetics by Computer, version 3.4. Report 132. Institute of Phonetic Sciences of the University of Amsterdam.

Bohas, G. (1997). *Matrices, etymons, raciness. Éléments d'une théorie lexicologique du vocabulaire arabe*. Leuven-Pars: Peeters.

Bohr, N. (1948). On the notion of causality and complementarity. *Dialectica* **2**, 312–319.

Bohr, N. (1958). *Atomic physics and human knowledge*. New York: Wiley.

Booij, G. (1977). *Dutch morphology*. *A study of word formation in Generative Grammar*. Lisse: The Peter de Ridder Press (reprinted Dordrecht: Foris Publications).

Booij, G. (1981). *Generatieve fonologie van het Nederlands*. Utrecht: Het Spectrum.

Booij, G. (1983). Principles and parameters in prosodic phonology. *Linguistics* **83**, 249–280.

Booij, G. (1984). Neutral vowels and the autosegmental analysis of Hungarian vowel harmony. *Linguistics* **22**, 629–641.

Booij, G. (1985). The interaction of phonology and morphology in prosodic phonology. *Studies in the interaction of phonology and morphology*, edited by E. Gussmann, 23–35. Lublin: Redakcja Wydawnictw.

Booij, G. (1997). Non-derivation phonology meets Lexical Phonology. *Derivations and constraints in phonology*, edited by I. Roca, 261–289. Oxford: Clarendon Press.

Booij, G. (1999). The role of the prosodic word in phonotactic generalizations. *Studies on the phonological word*, edited by T.A. Hall & U. Kleinhenz, 47–73. Amsterdam: John Benjamins Publishing Company.

Booij, G. (2000). Morpheme structure constraints and the phonotactics of Dutch. *The syllable*: *views and facts*, edited by H.G. van der Hulst & N.A. Ritter, 69–92. Berlin: Mouton de Gruyter.

Bopp, F. (1846). Über das Georgische in sprachverwandtschaftlicher Beziehung. *Philosophische und Historische Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin*. Berlin: Akademie der Wissenschaften.

Bopp, F. (1847). *Die kaukasischen Glieder des indoeuropäischen Sprachstammes*. Berlin.

Borowsky, T. (1987). Antigemination in English phonology. *Linguistic Inquiry* **18**, 671–678.

Borowsky, T. (1993). On the word level. *Phonetics and Phonology* **4**. *Studies in Lexical Phonology*, edited by S. Hargus & E.M. Kaisse, 199–232. San Diego: Academic Press.

Bouda, K. (1949). *Baskisch-kaukasische Etymologien*. Heidelberg: Winter.

Bouda, K. (1950). Beiträge zur etymologischen erforschung des Georgischen. *Lingua* **2**, 291–307.

Bouda, K. (1954–56). Südkaukasisch-nordkaukasiche Etymologien. *Die Welt des Orients* **II**. Göttingen.

Botha, R. (1971). *Methodological aspects of Transformational Generative Phonology*. The Hague: Mouton.

Brasington, R. (1997). Sound patterns: Phonotactics. Unpublished paper. University of Reading.

Broselow, E. (1991). The structure of fricative-stop onsets. Paper presented at the conference on *Phonological Feature Organization*. LSA Summer School. Santa Cruz, California.

Brosset, M. (1849). *Rapports sur une voyage archéologique dans la Géorgie en dans l'Arménie*. St. Petersburg.

Browman, C.P. & L. Goldstein (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica* **45**, 140–155.

Brown, G. (1969). Syllable and redundancy rules in generative phonology. *Journal of Linguistics* **5**, 1–17.

Bush, R. (1997). *Georgian syllable structure*. MA thesis, UC Santa Cruz.

Butskhrikidze, M. (1994). Implicational universals on relationship between the word stress and devoicing of final obstruents (in fixed stress languages). *Bulletin of Georgian Academy of Sciences* **149**, 334–336.

Butskhrikidze, M. (1995). The category of inalienable/alienable possession. (On the Kartvelian languages). Paper presented at the conference of young researchers at the Institute of Oriental Studies. (In Georgian).

Butskhrikidze, M. (1996). Morphological factors determining fixed accent placement. *Bulletin of the Georgian Academy of Sciences* **153**, 312–315.

Butskhrikidze, M. (1998a). *Piksirebuli makhvilis t'ip'ologia*. [Typology of fixed accent]. Candidate's dissertation, Tbilisi State University.

Butskhrikidze, M. (1998b). Phonetic evidence for the Georgian harmonic groups. Unpublished paper. Holland Institute of Generative Linguistics. Leiden.

Butskhrikidze, M. & V.J. van Heuven (2001). Georgian harmonic clusters as complex segments. A perceptual experiment. *Linguistics in the Netherlands* **18**, edited by T. van der Wouden & H. Broekhuis, 31–44. Amsterdam: John Benjamins Publishing Company.

Butskhrikidze, M. & J.M. van de Weijer (2001a). On v-metathesis in Modern Georgian. *Surface syllable structure and segment sequencing*, edited by E.V. Hume, N.S.H. Smith & J.M. van de Weijer. *HIL Occasional Papers* **4**, 91–101. Leiden: Holland Institute of Generative Linguistics.

Butskhrikidze, M. & J.M. van de Weijer (2001b). On de-affrication in Modern Georgian. *Linguistics in the Netherlands* **18**, edited by T. van der Wouden & H. Broekhuis, 45–55. Amsterdam: John Benjamins Publishing Company.

Bybee, J., R. Perkins & W. Pagliuca (1994). *The evolution of grammar*. Chicago: University of Chicago Press.

Campbell, L. (1974). Phonological features: problems and proposal. *Language* **50**, 52–66.

Carlson, K. (1997). Reduplication and sonority in Nakanai and Nuxalk. *Proceedings of the Fourteenth Eastern Student Conference on Linguistics '97*, edited by J. Austin & A. Lawson, 23–33. Ithaca: CLC Publications.

Charette, M. (1990). Licence to govern. *Phonology* **7**, 233–253.

Chikobava, A. (1942). *Sakhelis pudzis udzvelesi agebulebisatvis kartvelur enebshi*. [The oldest structure of noun bases in Kartvelian languages]. Tbilisi: Sakartvelos SSR Mecnierebata Ak'ademiis Gamomcemloba.

Chikobava, A. (1965). *Iberiul-k'avk'asiur enata shesc'avlis ist'oria*. [A history of the study of Ibero-Caucasian languages]. Tbilisi.

Chikobava, A. (1971). *Gruzinskij jazyk*. [The Georgian language]. Tbilisi: Ganatleba.

Chikobava, A. (1986). *Kartuli enis ganmart'ebiti leksik'oni.* [Explanatory dictionary of Georgian]. Tbilisi: Kartuli Sabch'ota Encik'lop'ediis Mtavari Samecniero Redakcia.

Chitoran, I. (1998). Georgian harmonic clusters: phonetic cues to phonological representation. *Phonology* **15**, 121–141.

Cho, Y.Y. & T. King. (1997). Semi-syllables and Georgian clusters. Handout presented at *10th Non-Slavic Languages Conference*, University of Chicago.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Massachusetts: MIT Press.

Chomsky, N. & M. Halle (1968). *The sound pattern of English*. New York: Harper & Row.

Chrétien, C.D. (1965). The statistical structure of the Proto-Austronesian morph. *Lingua* **14**, 243–270.

Chù, M.Ng. & V.Th.Th. Nguyệt (1997). *Vietnamese for everyone*. Selangor Darul Ehsan: Pelanduk Publications.

Clements, G.N. (1989). *A unified set of features for consonants and vowels*. Ms, Cornell University and University of Paris VII.

Clements, G.N. (1990). The role of the sonority cycle in core syllabification. *Papers in Laboratory Phonology* **1**. *Between the grammar and physics of speech*, edited by J. Kingston & M. Beckman, 283–333. Cambridge: Cambridge University Press.

Clements, G.N. & S.J. Keyser (1983). *CV phonology*. *A generative theory of the syllable*. Cambridge, Massachusetts: MIT Press.

Collinge, N. (1990). *An encyclopaedia of language*. London & New York: Routledge.

Contini-Morava, E. & Y. Tobin (1984). Introduction. *Between grammar and lexicon*, edited by E. Contini-Morava & Y. Tobin. Amsterdam: John Benjamins Publishing Company.

Cowgill, W. (1963). A search for universals in Indo-European diachronic morphology. *Universals of Language*, edited by J.H. Greenberg. 114–141. Cambridge, Massachusetts: MIT Press.

Davis, S. (1985). On the internal organization of syllable structure. *Working Papers in Linguistics from A–Z* **3**. University of Arizona.

Davis, S. (1997). Issues of syllable contact in Optimality Theory. Paper presented at UC Santa Cruz.

Deeters, G. (1926). Armenisch und Südkaukasisch. Ein Beitrag zur Frage der Sprachmischung. Caucasica, Fasc. III, Leipzig.

Deeters, G. (1930). *Das kharthwelische Verbum. Vergleichende Darstellung des Verbalbaus der südkaukasischen Sprachen.* Leipzig.

Dell, F. & M. Elmedlaoui (1985). Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber. *Journal of African Languages and Linguistics* **7**, 105–130.

Deprez, V. (1988). Georgian complex segments. *North Eastern Linguistic Society* **18**, 109–123.

Derwing, B.L. (1980). Against autonomous linguistics. *Evidence and argument in linguistics*, edited by Th.A. Perry, 163–190. Berlin: Walter de Gruyter.

Diakonoff, I.M. (1970). Problems of root-structure in Proto-Semitic. *Archiv Orientálnì* **38**, 453–480.

Diakonoff, I.M. (1988). *Afrasian languages*. Moscow: Nauka.

Đình-hoà, Ng (1997). *Vietnamese*. Amsterdam: John Benjamins Publishing Company.

Dirks, S. (1953). Campa (Arawak) phonemes. *International Journal of American Linguistics* **19**, 302–304.

Dixon, R.M.W. (1970). Olgolo syllable structure and what they are doing about it. *Linguistic Inquiry* **1**, 273–276.

Dixon, R.M.W. (1977). *A grammar of Yidiɲ.* Cambridge: Cambridge University Press.

Dixon, R.M.W. (1980). *The languages of Australia*. Cambridge: Cambridge University Press.

Downing, L. (1999). Prosodic stem ≠ prosodic word in Bantu. *Studies on the Phonological Word*, edited by T.A. Hall & U. Kleinhenz, 73–99. Amsterdam: John Benjamins Publishing Company.

Dzidzishvili, M. (1966). Tankhmovanta k'omp'leksisa da ponet'ik'uri p'rocesis urtiertmimartebis shesakheb kartulshi. [The relationship between the consonant complexes and phonetic processes in Georgian]. *Iberiul-K'avk'asiuri Enatmecniereba* **15**, 228–232.

Edwards, W.F. (1978). A preliminary sketch of Arekuna (Carib) phonology. *International Journal of American Linguistics* **44**, 223–227.

Eikhenvald, A.Y. (1996). Words, phrases, pauses and boundaries. Evidence from South American Indian languages. *Studies in Language* **20**, 487–517.

Ertelishvili, P. (1970). *Zmnuri pudzeebis ponemat'uri st'rukt'urisa da ist'oriis sak'itkhebi kartulshi*. [Phonemic structure and historical aspects of verbal stems in Georgian]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.

Ertelishvili, P. (1980). *Sakhelur pudzeta ponemat'uri st'rukt'urisa da ist'oriis sak'itkhebi*. [Phonemic structure and historical aspects of nominal stems in Georgian]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.

Ewen, C.J. & H.G. van der Hulst (2001). *The phonological structure of words. An introduction*. Cambridge: Cambridge University Press.

Fähnrich, H. (1978). *Enata genet'uri natesaobis k'rit'eriumebi da adrindeli saertokartveluri pudze enis dzireul morpemata st'rukt'uris zogierti sak'itkhi*. [Phonemic structure and historical aspects of nominal stems in Georgian. Criteria of the genetic affiliation of languages and some problems of the structure of root morphemes in Proto-Kartvelian]. Tbilisi.

Fischer-Jørgensen, E. (1952). On the definition of phoneme categories on a distributional basis. *Acta Linguistica* **7**, 8–39. Copenhagen: Fascicule 1–2 Einer Munksgaard.

Fodor, J.A & J.J. Katz (1964). *The structure of language*. *Readings in the philosophy of language*. New Jersey: Prentice-Hall, Inc.

Foley, J. (1970). Phonological distinctive features. *Folia Linguistica* **4**, 87–92.

Fowler, M. (1954). The segmental phonemes of Sanskritized Tamil. *Language* **30**, 360–367.

Fowler, C.A. (1983). Converting sources of evidence on spoken and perceived rhythms of speech. Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology* **112**, 386–412.

Fowler, C.A. & L. Tassinary (1981). Natural measurement criteria for speech. The anisochrony illusion. *Attention and performance* **IX**, edited by J. Long & A. Baddeley, 521–536. Hillsdale, New Jersey: Erlbaum.

Fromkin, V.A. (1968). Speculations on performance models. *Journal of Linguistics* **4**, 47–68.

Fromkin, V.A. (1971). The non-linear nature of anomalous utterances. *Language* **47**, 27–52.

Fudge, E. (1969). Syllables. *Journal of Linguistics* **5**, 253–287.

Gamkrelidze, Th. (1974). The problem of 'l'arbitraire du signe'. *Language* **50**, 102–110.

Gamkrelidze, Th. (1989). *Alphabetic writing and the Old Georgian script.* Tbilisi: Publishing House of the Tbilisi State University.

Gamkrelidze Th. & G. Machavariani (1965). *Sonant'ta sist'ema da ablaut'i Kartvelur enebshi*. [Sonant system and ablaut in Kartvelian languages]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.

Gardthausen, V. (1879). *Griechische Palaeographie*. Leipzig: Teubner.

Goedemans, R.W.N. (1998). *Weightless segments. A phonetic and phonological study concerning the metrical irrelevance of syllable onsets*. LOT dissertation series **9**, Utrecht: Landelijke Onderzoeksschool Taalwetenschap.

Granberry, J. (1956). Timicua I. Prosodics and phonemics of the Mocama dialect. *International Journal of American Linguistics* **22**, 97–105.

Greenberg, J.H. (1950). The patterning of root morphemes in Semitic. *Journal of the Linguistic Circle of New York* **6**, 162–181.

Greenberg, J.H. (1960). A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* **26**, 178–194.

Greenberg, J.H. (1970). The patterning of root morphemes in Semitic. *Word* **6**, 162–181.

Greenberg, J.H. (1978). Some generalizations concerning initial and final consonant clusters. *Universals of human language* **2**, *Phonology* edited by J.H. Greenberg, 243–280. Stanford: Stanford University Press.

Gudava, T. (1979). Bagismieri tankhmovnebi chkamierta c'in Megrulshi. [Bilabial consonants in front of obstruents in Megrelian]. *Saenamecniero K'rebuli*. Tbilisi.

Güldenstädt, J.A. (1787). *Reisen durch Russland und caucasischen Gebürge*. Bd. 2. Petersburg.

Gussmann, E. (1992). Resyllabification and delinking: the case of Polish voicing. *Linguistic Inquiry* **23**, 25–56.

Gvinadze, C. (1970). Tavk'idura tankhmovantk'omp'leksebi kartulshi. [Word-initial consonant sequences in Georgian]. *Student'ta Samecniero K'onperencia XXXII*, 21–24. Mokhsenebata tezisebi. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.

Hála, B. (1961). La syllabe, sa nature, son origine et ses transformations. *Orbis* **19**, 69–143.

Hall, T.A. (1999). The phonological word: a review. *Studies on the phonological word*, edited by T.A. Hall & U. Kleinhenz, 1–23. Amsterdam: John Benjamins Publishing Company.

Halle, M. & J.-R. Vergnaud (1978). *An essay on stress*. Cambridge, Massachusetts: MIT Press.

Harris, J. (1994). *English sound structure*. Oxford: Blackwell.

Harris, J. & E. Gussmann (1998). Final codas: why the west was wrong. *Structure and interpretation*, edited by E. Cyran, 139–162. Lublin: Wydawnictwo Folium.

Harris, R. (1987). Language as social interaction: integrationalism vs. segregationalism. *Language Sciences* **9**, 131–143.
Harris, R. (1990). *The foundations of linguistic theory: selected writings of Roy Harris*, edited by N. Love. London: Routledge.
Haugen, E. (1956). Syllabification in Kutenai. *Language* **22**, 196–201.
Hayes, B.P. (1981). *A metrical theory of stress rules*. Doctoral dissertation. MIT.
Hayes, B.P. (1989). The prosodic hierarchy in meter. *Rhythm and meter*, edited by P. Kiparsky & G. Youmans, 201–260. Orlando, Florida: Academic Press.
Hayes, B.P. (1995). *Metrical stress theory: principles and case studies.* Chicago: University of Chicago Press.
Hjelmslev, L. (1936). On the principles of phonematics. *Proceedings of the Second International Congress of Phonetic Sciences*, 49–54. Cambridge: Cambridge University Press.
Hjelmslev, L. (1939). The syllable as a structural unit. *Proceedings of the Third International Congress of Phonetic Sciences*, 266–272. Ghent: University of Ghent.
Hooper, J.B. (1972). The syllable in phonological theory. *Language* **48**, 525–540.
Hooper, J.B. (1976). *An introduction to Natural Generative Phonology.* New York: Academic Press.
Hulst, H.G. van der (1984). *Syllable structure and stress in Dutch*. Dordrecht: Foris.
Hulst, H.G. van der & J.M. van de Weijer (1985). Vowel harmony. *The handbook of phonological theory*, edited by J.A. Goldsmith, 495–534. Oxford: Blackwell.
Humboldt, W. von (1922). Über das Entstehen der grammatischen Formen, und ihren Einfluss auf die Ideen-Entwicklung. Berlin: Königliche Akademie der Wissenschaften.
Hyman, L. (1975). *Phonology. Theory and analysis*. New York: Holt, Rinehart and Winston.
Hyman, L. (1978). Word demarcation. *Universals of human language* **2**, *Phonology* edited by J.H. Greenberg, 443–470. Stanford: Stanford University Press.
Hyman, L. (1985). *A theory of phonological weight*. Dordrecht: Foris.
Hyman, L. (1987). Prosodic domains in Kukuya. *Natural Language and Linguistic Theory* **5**, 311–333.
Inkelas, S. (1989). *Prosodic constituency in the lexicon*. Doctoral dissertation, Stanford University.
Israel, J. (1979). *The language of dialectics and the dialectics of language*. London: Harvester Press.
Itô, J. (1986). *Syllable theory in Prosodic Phonology*. Doctoral dissertation, University of Massachusetts, Amherst. Published in 1988, New York, Garland.
Itô, J. (1989). A prosodic theory of epenthesis. *Natural Language and Linguistic Theory* **7**, 217–259.
Itô, J. & R.A. Mester (1986). The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* **17**, 49–74.
Itô, J. & R.A. Mester (1992). Weak layering and word binarity. *Linguistic Research Council Working Papers*, Santa Cruz, California.

Itô, J. & R.A. Mester (1995). The core-periphery structure of the lexicon and constraints on ranking. *University of Massachusetts Occasional Papers* **19**, edited by J. Beckman et al., 191–209. University of Massachusetts, Amherst.

Jacobson, S. (1985). Siberian Yupik and Central Yupik prosody. *Yupic Eskimo prosodic systems: descriptions and comparative systems*, edited by M. Krauss, 25–46. Fairbanks, Alaska: Alaska Native Language Centre.

Jakobson, R. (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Uppsala Universitets Arsskrift.

Jakobson, R. (1949). The phonetic and grammatical aspects of language and their interrelation. *Proceedings of the Sixth International Congress of Linguistics*, 103–114. Paris: Klincksiek.

Jakobson, R. (1971). *Selected writings*, vol. 2: *Word and language*. The Hague: Mouton.

Jakobson, R. & L. Waugh (1979). *The sound shape of language*. Brighton: Harvester Press.

Jespersen, O. (1904). *Lehrbuch der Phonetik*. Leipzig: Teubner.

Jespersen, O. (1924). *The philosophy of grammar*. New York: Holt.

Kahn, D. (1976). *Syllable-based generalizations in English phonology*. Doctoral dissertation, MIT.

Kasevič, B.V. (1986). *Morphonologija.* [Morphonology]. Leningrad: Izdatel'stvo Leningradskogo Universiteta.

Kasevič, B.V. & S.E. Jaxontov (1982). *Kvantitativnaja tipologija jazykov Azii i Afriki.* [A qualitative typology of the languages of Asia and Africa]. Leningrad: Izdatel'stvo Leningradskogo Universiteta.

Kaye, J., J. Lowenstamm & J.-R. Vergnaud (1985). The internal structure of phonological elements: A theory of charm and government. *Phonology Yearbook* **2**, 303–326.

Kaye, J., J. Lowenstamm & J.-R. Vergnaud (1990). Constituent structure and government in phonology. *Phonology* **7**, 193–231.

Kehrein, W. (1999). Phonology without affricates. Presented at *HIL Phonology Conference* **4**, Leiden University.

K'ek'elidze, K. (1929). Kartuli k'ult'uris ori dgesasc'auli. [Two celebrations of the Georgian culture]. *Mnatobi* **5**–**6**, 151–167; **7**, 135–150.

Kempen, G. & P. Huijbers (1983). The lexicalization process in sentence production and naming. Indirect selection of words. *Cognition* **14**, 185–209.

Kenstowicz, M. (1994). *Phonology in generative grammar*. Oxford: Blackwell.

Key, H. & M. Key (1953). The phonemes of Sierra Nahua. *International Journal of American Linguistics* **19**, 53–56.

Khundadze, S. (1901). *Salit'erat'uro kartuli*. [Literary Georgian]. Kutaisi.

K'iziria, N. (1985). Khmovanta dist'anciuri gavlenis shesakheb kartulshi. *Iberiul-K'avk'asiuri Enatmecnierebis C'elic'deuli* **20**, 91–98.

Kobalava, I. (1967). CVC t'ip'is dzirta ponemuri st'rukt'urisatvis kartulshi. [The phonemic structure of roots of the CVC type in Georgian]. *Tanamedrove Zogadi da Matemat'ik'uri Enatmecnierebis Sak'itkhebi* **2**, 183–192. Tbilisi: Mecniereba.

Kohler, K.J. (1966). Is the syllable a phonological universal? *Journal of Linguistics* **2**, 207–208.

Krámský, J. (1964). A quantitative phonemic analysis of Italian mono-, di-, trisyllabic words. *Travaux Linguistiques de Prague* **1**, 1290–1243.

Krupa, V. (1967). Dissociation of like consonants. *Asian and African Studies* **3**, 37–44.

Krupa, V. (1971). The phonotactic structure of the morph in Polynesian languages. *Language* **47**, 668–684.

Kuryłowicz, J. (1948). Contribution à la théorie de la syllabe. *Biuletyn Polskiego Towarzystwa Językoznawczego* **8**, 80–114.

Kuryłowicz , J. (1975). *Esquisses linguistiques II*. München, Wilhelm Fink Verlag.

Kutelia, N. (1956). *Dve sistemy garmoničnyx soglasnyx v Gruzinskom jazyke* (Eksperimental'noe issledovanie). [Two systems of harmonic clusters in Georgian (experimental studies)]. Doctoral dissertation, Tbilisi State University.

Leben, W. (1973). *Suprasegmental phonology*. Doctoral dissertation, MIT.

Leben, W. (1978). The representation of tone. *Tone: a linguistic survey*, edited by V. Fromkin, 177–220. New York: Academic Press.

Lehmann, Chr. (1986). Grammaticalization and linguistic typology. *General Linguistics* **26**, 3–23.

Levelt, C.C. (1994). *On the acquisition of place.* HIL dissertations series 8. The Hague: Holland Academic Graphics.

Levelt, W.J.M. (1992). Accessing words in speech production: stages, processes and representations. *Cognition* **42**, 1–22.

Levelt, W.J.M. & B. Maassen (1981). Lexical search and order of mention in sentence production. *Crossing the boundaries in Linguistics. Studies presented to Manfred Bierwisch*, edited by W. Klein & W.J.M. Levelt, 221–252. Dordrecht: Reidel.

Levin, J. (1985). *A metrical theory of syllable.* Doctoral dissertation, MIT.

Li, C.N. & S.A. Thompson (1987). Chinese. *The world's major languages*, edited by B. Comrie, 811–833. London: Croom Helm.

Liberman, M. & A. Prince (1977). On stress and linguistic rhythm. *Linguistic Inquiry* **8**, 249–336.

Lit, J. van (1988). *Clusters en segmenten in het Georgisch.* [Clusters and segments in Georgian]. MA Thesis, Leiden University.

Lowenstamm, J. (1996). CV as the only syllable type. *Current trends in phonology: models and methods*, edited by J. Durand & B. Laks, 419–442. Manchester: University of Salford Publications.

Machavariani, G. (1957). On the accent in Kartvelian languages. Paper given at the meeting 'Linguistic Discussions', Tbilisi State University. Published in the article: 'Annotated bibliography of the writings by G. Machavariani'. *Mimomkhilveli* **9**. Tbilisi, 1972.

Machavariani, M. (1965). *Saerto Kartveluri k'onsonant'uri sist'ema.* [The Common-Kartvelian consonant system]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.

Mackenzie, D.N. (1967). Notes on the transcription of Pahlavi. *Bulletin of the School of Oriental and African Studies* **30**, 17–29.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Magnusson, W.L. (1967). Complementary distribution among the root patterns of Proto-Indo-European. *Linguistics* **34**, 17–25.

Magnusson, W.L. (1979). Chronostructural orientation of root pattern development. *Linguistics* **17**, 1055–1056.

Marr, N. (1908). *Osnovnye tablicy k grammatike drevnegruzinskogo jazyka s predvoritel'nym soobščeniem o rodstve gruzinskogo jazyka s semitičeskimi.* [Foundations of Old Georgian grammar. A tentative proposal on the genetic affiliation of Georgian to the Semitic languages]. St. Petersburg.

Marr, N. (1925). *Grammatika drevneliteraturnogo gruzinskogo jazyka*. [A grammar of Old Literary Georgian]. *Materialy po Jafeticeskomy Jazykoznaniju* **12**. Leningrad.

Marslen-Wilson, W.D. & P. Zwitserlood (1989). Accessing spoken words: the importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance* **22**, 576–585.

Mathesius, V. (1929). La structure phonologique du lexique de tchèque moderne. *Travaux du Cercle Linguistique de Prague* **1**, 148–152.

McCarthy, J. (1986). OCP effects: gemination and antigemination. *Linguistic Inquiry* **17**, 207–264.

McCarthy, J. & A. Prince (1986). *Prosodic morphology*. Ms, University of Massachusetts, Amherst & Brandeis University.

McCarthy, J. & A. Prince (1993). *Prosodic Morphology*. Ms, University of Massachusetts, Amherst & Rutgers University.

McCarthy, J. & A. Prince (1994). The emergence of the unmarked: Optimality in Prosodic Morphology. *Proceedings of the North East Linguistic Society* **24**, edited by M. Gonzalez. Graduate Linguistic Student Association, University of Massachusetts, Amherst.

McCarthy, J. & A. Prince (1995). Faithfulness and reduplicative identity. *University of Massachusetts Occasional Papers in Linguistics* **18**, edited by J. Beckman et al., 249–384. University of Massachusetts, Amherst.

McClelland, J.L. & J.L. Elman (1986). The TRACE model of speech perception. *Cognitive Psychology* **18**, 1–86.

Mdivani, N. (1983). *Tankhmovanta shetavsebadobis sak'itkhebi khundzur enashi*. [The aspects of consonant combinatorics in Avarian]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.

Meijer, P.J.A. (1994). *Phonological encoding. The role of suprasegmental structures*. Doctoral dissertation, Nijmegen University.

Meijer, P.J.A. (1996). Suprasegmental structures in phonological encoding. The CV structure. *Journal of Memory and Language* **35**, 840–853.

Mejlax, M. (1973). Indeksy morfologičeskoj tipologii. [Indices of morphological typology]. *Problemy grammatičeskoj modelirovanija*, edited by A.A. Zaliznjak, 155–170. Moscow: Nauka.

Melikishvili, I. (1976). *Mark'irebis mimarteba ponologiashi*. [Markedness relations in phonology]. Tbilisi: Mecniereba.

Melikishvili, I. (1979). Marcvlis agebis zogadi p'rincip'i da khshulta ganac'ileba CVC t'ip'is dzirebshi saerto kartvelursa da saerto indoevrop'ulshi. [A general principle of syllable construction and the distribution of stops in roots of the CVC type in Proto Kartvelian and Proto Indo-European]. *Macne, Enisa da Lit'erat'uris Seria* **4**, 132–144. Tbilisi: Mecniereba.

Melikishvili, I. (1997). *Saerto-kartveluri dziri st'rukt'uruli da t'ip'ologiuri tvalsazrisit*. [The Proto Kartvelian root in structural and typological aspects]. Doctoral dissertation, Tbilisi State University.

Mel'nikov, G.P. (1966). *Sistemnyij analiz pričin svoeobrazija semitskogo konsonantizma*. [Systemic analysis of reasons for peculiarities of the Semitic consonants]. Moscow: Nauka.

Milewski, T.Z. (1967). *Językoznawstwo*. [Linguistics]. Warsaw: Państwowe Wydawnictwo Naukowe.

Milewski, T.Z. (1969). *Z zagadnien językoznawstwa ogólnego i historycznego*. Warsaw: Państwowe Wydawnictwo Naukowe.

Mohanan, K.P. (1995). The organization of the grammar. *The handbook of phonological theory*, edited by J.A. Goldsmith, 24–69. Oxford: Blackwell.

Mous, M. (1993). *A grammar of Iraqw*. Hamburg: Buske.

Murray, R. & T. Vennemann (1983). Sound change and syllable structure in Germanic phonology. *Language* **59**, 514–528.

Myers, S. (1987). *Tone and the structure of words in Shona*. Doctoral dissertation, University of Massachusetts, Amherst. Published in 1991, New York: Garland.

Nebieridze, G. (1975). *Salit'erat'uro kartuli enis Generat'iul-ponologiuri modeli da misi agebis p'rincip'ebi*. [A Generative phonological model of literary Georgian and the principles of its construction]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.

Nepveu, D. (1994). *Georgian and Bella Coola: Headless syllables and syllabic obstruents*. MA thesis, UC Santa Cruz.

Nespor, M. & I. Vogel (1986). *Prosodic phonology*. Dordrecht: Foris.

Nida, E.A. (1949). *Morphology. The descriptive analysis of words*. Ann Arbor: University of Michigan Press.

Norris, D. & A. Cutler (1988). The relative accessibility of phonemes and syllables. *Perception and Psycholinguistics* **46**, 541–550.

O'Connor, J.D. & J.L.M. Trim (1953). Vowel, consonant, and syllable. A phonological definition. *Word* **9**, 103–122.

Odendal, F.F. (1962). *Die structuur van afrikaanse wortelmorfeem* [The structure of Afrikaans root morphemes]. Cape Town: H.A.U.M.

Oranskij, I.M. (1960). *Vvedenie v iranskuju filologiju*. [An introduction to Iranian philology]. Moscow: Podred. AA Frejmana.

Paradis, C. & J.-P. Prunet (1993). On the validity of Morpheme Structure Constraints. *Canadian Journal of Linguistics* **38**, 235–256.

Peperkamp, S. (1997). *Prosodic words*. HIL dissertation series **34**. The Hague: Holland Academic Graphics.

Pierce, C.S. (1931). *Collected papers.* Cambridge, Massachusetts: Harvard University Press.

Pike, K.L. (1947). *Phonetics: A technique for reducing language to writing*. Ann Arbor: University of Michigan Press.

Pike, K. & E. Pike (1947). Immediate constituents of Mazatec syllables. *International Journal of American Linguistics* **13**, 78–91.

Pike, E.V. & E. Scott (1962). The phonological hierarchy of Marinahua. *Phonetica* **8**, 1–8.

Piotrowski, M., I. Roca & A. Spencer (1992). Polish yers and lexical syllabicity. *The Linguistic Review* **9**, 27–67.

Pisowicz, A. (1985). *Origins of the New and Middle Persian phonological systems*. Krakow: Uniwersytet Jagiełłonski, Rozprawy Habilitacyjne.

Pitt, M.A. & J. McQueen (1998). Is compensation for coarticulation mediated by he lexicon? *Journal of Memory and Language* **39**, 347–370.

Popova, E.N. (1989). *Sootnošenije morphemy i sloga v russkom jazyke*. [The correlation between the morpheme and the syllable in Russian]. Doctoral dissertation, Moscow University.

Popper, K. (1963). *Conjectures and refutations*. London: Routledge & Kegan Paul.

Postal, P. (1968). *Aspects of phonological theory*. New York: Harper & Row.

Prince, A. & P. Smolensky (1993). *Optimality Theory*. Ms, University of Massachusetts, Amherst & Rutgers University.

Pulgram, E. (1970). *Syllable, word, nexus, cursus*. The Hague: Mouton.

Pulleyblank, E.G. (1989). The role of coronal in articulator-based features. *Papers from the 25th Annual Regional Meeting of the Chicago Linguistic Society*, edited by C. Wiltshire, R. Graczyk & B. Music, 379–393. Chicago: Chicago Linguistic Society.

Raffelsiefen, R. (1999). Diagnostics for prosodic words revisited: the case of historically prefixed words in English. *Studies on the phonological word*, edited by T.A. Hall & U. Kleinhenz, 133–203. Amsterdam: John Benjamins Publishing Company.

Reighard, J. (1972). Labiality and velarity in consonants and vowels. *Papers from the 8th Annual Regional Meeting of the Chicago Linguistic Society*, edited by P.M. Peranteau, J.N. Levi & G.C. Phares, 533–543. Chicago: Chicago Linguistic Society.

Revithiadou, A. (1999). *Headmost accent wins. Head dominance and ideal prosodic form in lexical accent systems*. LOT dissertation series **15**. Utrecht: Landelijke Onderzoeksschool Taalwetenschap.

Rhee, S.J. (2002). *Empty nuclei in Korean*. LOT dissertation series **61**. Utrecht: Landelijke Onderzoeksschool Taalwetenschap.

Ringen, J.D. (1980). Linguistic facts. *Evidence and argument in linguistics*, edited by T.A. Perry, 97–132. Berlin: Walter de Gruyter.

Robins, R.H. & N. Waterson (1952). Notes on the phonetics of the Georgian word. *Bulletin of the School of Oriental and African Studies* **14**, 52–72.

Roelofs, A. (1996). Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language* **35**, 854–876.

Roelofs, A. & A.S. Meyer (in press). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology*: *Learning, Memory and Cognition*.

Rosetti, A. (1962). La syllabe phonologique. *Proceedings of the Fourth International Congress of Phonetic Sciences* **4**, 490–499. The Hague: Mouton.

Russell, K. (1999). The "word" in two polysynthetic languages. *Studies on the phonological word*, edited by T.A. Hall & U. Kleinhenz, 203–222. Amsterdam: John Benjamins Publishing Company.

Sapir, E. (1921). *Language: Introduction to the Study of Speech*. New York: Harcourt, Brace & World, Inc.

Saporta, S. (1955). Frequency of consonant clusters. *Language* **33**, 25–31.

Saporta, S. (1963). Phoneme distribution and language universals. *Universals of Language*, edited by J.H. Greenberg, 61–67. Cambridge, Massachusetts: MIT Press.

Sardschweladse, Z. & H. Fähnrich (1990). *Kartvelur enata et'imologiuri leksik'oni*. [Etymological dictionary of Kartvelian languages]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.

Saussure, F. de (1916). *Cours de linguistique générale.* Lausanne & Paris: Payot.

Sawusch, J.R. & P.W. Jusczyk (1981). Adaptation and contrast in the perception of voicing. *Journal of Experimental Psychology: Human Perception and Performance* **7**, 408–421.

Schane, S.A. (1968). *French phonology and morphology*. Cambridge, Massachusetts: MIT Press.

Schiller, N.O. (1997). *The role of the syllable in speech production. MPI Series in Psycholinguistics* **2**. Nijmegen University.

Schlegel, A.W. von (1918). *Observations sur la langue et la literature provençales.* Paris: Librarie Grecque-latine-allemande.

Schönefeld, D. (1997). The place of the lexicon in linguistic models – how natural are descriptions of the language system? *Lexicology* **3**, 1–60.

Selkirk, E.O. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry* **11**, 563–605.

Selkirk, E.O. (1982). Syllables. *The structure of phonological representations*, edited by H.G. van der Hulst & N. Smith, vol. 2, 337–383. Dordrecht: Foris.

Selkirk, E.O. (1984a). On the major class features and syllable theory. *Language sound structure*, edited by M. Aronoff & R. Oehrle, 107–136. Cambridge, Massachusetts: MIT Press.

Selkirk, E.O. (1984b). *Phonology and syntax. The relation between sound and structure*. Cambridge, Massachusetts: MIT Press.

Shanidze, A. (1973). *Kartuli gramat'ik'is sapudzvlebi*. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.

Shaw, P. (1993). The prosodic constituency of minor syllables. *Proceedings of the Annual Meeting of the West Coast Conference on Formal Linguistics* **12**, 117–132.

Shibatani, M. (1973). Where morphology and syntax clash. A case study of Japanese aspectual verbs. *Gengo Kenkyū* **64**, 65–96.

Sgall, P. (1986). Classical typology and modern linguistics. *Folia Linguistica* **20**, 15–28.

Sigurd, B. (1955). Rank order of consonants established by distributional criteria. *Studia Linguistica* **9**, 8–20.

Sigurd, B. (1965). *Phonotactic structure in Swedish*. Lund: Uniskol.

Skalička, V. (1970). Die Typologie der Lautsysteme. *Proceedings of the International Congress of Phonetic Sciences* **6**. Prague.

Skalička, V. (1979). *Typologische Studien*. Braunschweig: Vieweg.

Spencer, A. (1996). *Phonology: theory and description*. Oxford: Blackwell.

Stebnickij, S.N. (1934). *Itel'menskij jazyk. Jazyki i pismennost' narodov Severa*, vol. 3. [The Itelmen language. Languages and writing of the people of North]. Moscow.

Steriade, D. (1989). Affricates are stops. Paper presented at the MIT conference on Features, MIT.

Stetson, R.H. (1951). *Motor phonetics. Study of speech movements in action*. 2nd edn. Amsterdam: North-Holland Publishing Company.

Stubbs, M. (1996). *Text and corpus analysis*. Oxford: Blackwell.

Tevdoradze, I. (1978). *Kartuli p'rosodiis sak'itkhebi*. [Issues in Georgian prosody]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.

Thornton, A.M. (1996). On some phenomena of prosodic morphology in Italian: *accordiamenti*, hypocoristics and prosodic delimitation. *Probus* **6**, 81–112.

Toft, Z. (1999). Grunts and gutturals in Georgian. An investigation into initial consonant clusters. *SOAS Working Papers in Linguistics* **9**, 275–297.

Toolan, M. (1999). Integrational linguistics, relevance theory and stylistic explanation: a reply to MacMahon. *Language and Literature* **8**, 255–268.

Torre, E.J. van der (2000). /pr/o/bl/ems in Dutch and Icelandic. Paper given at the *Conference of the Student Organization of Linguistics in Europe 2*. Lund.

Trnka, B. (1936). General laws of phonemic combinations. *Travaux du Cercle Linguistique de Prague* **6**.

Trubetzkoy, N.S. (1939). *Principles of phonology* (1st edn. 1939, translated from German by C.A.M. Baltaxe). Berkeley: University of California Press, 1969.

Tumanjan, E.G. (1971). *Drevnearmjanskij jazyk*. [The Old Armenian language]. Moscow.

Twaddell, W.F. (1939). Combinations of consonants in stressed syllables in German. *Acta Linguistica* **1**, 189–199.

Twaddell, W.F. (1941). Combinations of consonants in stressed syllables in German. *Acta Linguistica* **2**, 31–50.

Uturgaidze, T. (1976). *Kartuli enis ponemat'uri st'rukt'ura*. [The phonemic structure of the Georgian language]. Tbilisi: Mecniereba.

Uturgaidze, T. (1971). Harmoniul k'omp'leksta diakroniuli int'erpret'aciisatvis. [On the diachronic interpretation of harmonic clusters]. *Macne. Enisa da Lit'era'turis Seria* **4**, 111–125.

Vaux, B. (1998). *The phonology of Armenian*. Oxford: Clarendon Press.

Vennemann, T. (1971). The phonology of Gothic vowels. *Language* **47**, 90–132.

Vennemann, T. (1972). On the theory of syllabic phonology. *Linguistische Berichte* **18**, 1–18.

Vennemann, T. (1978). Universal syllabic phonology. *Theoretical Linguistics* **5**, 175–215.

Vijver, R. van de (1998). *The iambic issue*. HIL dissertation series **37**. The Hague: Holland Academic Graphics.

Vitevich, M.S. & P.A. Luce (1998). When words compete: levels of processing in spoken word recognition. *Psychological Science* **9**, 325–329.

Vitevich, M.S. & P.A. Luce (1999). Probabilistic phonotactics and neighbourhood activation in spoken word recognition. *Journal of Memory and Language* **40**, 374–408.

Vogt, H. (1958). Structure phonémique du Géorgien. *Norsk Tidsskrift for Sprogvidenskap* **18**, 5–90.

Vogt, H. (1961). *Kartuli enis ponemat'uri st'rukt'ura*. [The phonematic structure of the Georgian Language]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.

Waugh, L.R. (1978). Some remarks on the nature of the linguistic sign. *Sign, system and function. Papers of the First and Second Polish-American Semiotics Colloquia*, edited by J. Pelc, Th.A. Sebeok et al. Berlin: Mouton.

Weijer, J.M. van de (1996). *Segmental structure and complex segments*. Tübingen: Niemeyer.

Wurzel, W.U. (1989). *Flexionsmorphologie und Natürlichkeit. Ein Beitrag zur morphologischen Theoriebildung*. Berlin: Akademie Verlag.

Yip, M. (1988). The OCP and phonological rules: a loss of identity. *Linguistic Inquiry* **19**, 65–100.

Yip, M. (1994). Morpheme-level features: Chaoyang syllable structure and nasalization. Ms, University of California, Irvine. Available on the Rutgers Optimality Archive, http://roa.rutgers.edu [ROA-81].

Žgent'i, S. (1956). *Kartuli enis ponet'ik'a*. [A phonetic study of the Georgian language]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.

Žgent'i, S. (1965). Harmonische Komplexe von Konsonanten in der Khartvelischen Sprachen. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* **18**, 143–148.

Zubkova, L.G. (1988). Zvukovaja forma značaščix edinic jazyka i strukturnije xaralteristiki fonem. *Serija Literatury i Jazyka* **47**.

Zubkova, L.G. (1990). *Fonologičeskaija typologija slova.* [The phonological typology of the word]. Moscow: Izdatel'stvo Universiteta Druzhby Narodov.

Samenvatting (Summary in Dutch)

Het centrale onderwerp van deze dissertatie is het onderzoek van consonantsequenties, d.w.z. opeenvolgingen van consonanten van het type CCC. Uitgaande van de veronderstelling dat taal een open systeem is en vooruitlopend op het idee van relativiteit en complementariteit, wordt aangenomen dat er een relatie bestaat tussen vormen van het type CCC en CVCVCV. Meer specifiek wordt voorgesteld dat alle welgevormde consonantsequenties kunnen worden afgeleid van structuren van het type CVC, die het lexicale gedeelte van het woord, de stam, vormen. Wat betreft de representatie wordt een fonologische hiërarchie voorgesteld waarin de stam als domein zich bevindt tussen het segment- en het woorddomein. Dit maakt het mogelijk een verklaring te geven voor de fonotactische beperkingen op consonanten. De formele, semantische en structurele aspekten van de taal worden samen in één driedimensionale benadering voorgesteld om de fonotaxis op het woordniveau te verklaren. In die representatie is de lineaire sequentie van segmenten in een woord het vorm-gerelateerde aspect, morfologische structuur is het betekenis-aspect en fonologische principes zijn het structuur-aspect van het woord. Tot die fonologische principes behoren het Sonority Sequencing Principle, het Obligatory Contour Principle en andere principes, die opgevat werden als verschillende verschijningsvormen van een en hetzelfde fenomeen, het Balancing Principle. Het grootste gedeelte van deze dissertatie bestaat uit het testen van de voorspellingen die een dergelijk model doet en de hypotheses die aan de hand daarvan opgesteld werden in hoofdstukken 1 en 2.

De consonantsequenties van het Georgisch (bijv. /prckvn/, /mc'vrtn/, /brt'χ'/ in woorden zoals /prckvna/ 'pellen', /mc'vrtneli/ 'trainer', /brt'χ'eli/ 'plat') lijken door hun lengte en structuur een probleem te vormen voor onderzoekers die aannemen dat taal een systeem is. De fonotaxis en het gedrag van enkelvoudige consonanten en consonanten in clusters worden hier bestudeerd van de synchrone, de historische, en de comparatieve kant. In hoofdstuk 3 wordt vastgesteld dat er drie typen consonantsequenties in het Georgisch zijn die als complexe segmenten kunnen worden beschouwd: (i) zgn. harmonische groepen, (ii) C + /v/ en (iii) /s/ + obstruent. Het minimale woord in het Georgisch is bisyllabisch en sequenties van consonanten kunnen alleen in het stemdomein verschijnen. De hypothese dat consonantsequenties afgeleid zijn is getest door de patronen van combinaties van adjacente consonanten en consonanten die gescheiden door een vocaal in de stam voorkomen te vergelijken. Er werd een verrassende overeenkomst tussen beide gevonden, wat de voorspelling bevestigt dat clusters in het stamdomein maximaal tweeledig zijn. Deze clusters zijn het resultaat van deletie van de stamvocaal, een veel voorkomend proces zowel in het Oud Georgisch als in de moderne taal. Tevens wordt aangetoond

dat consonantale complexiteit in het Modern Georgisch een resultaat is van morfologische complexiteit, en ook van vocaaldeletie en de vorming van complexe segmenten.

In hoofdstuk 4 worden meer bewijzen gezocht om de claims uit hoofdstuk 3 te ondersteunen. Een perceptuele techniek werd gebruikt om de aanname dat harmonische groepen complexe segmenten zijn, te onderzoeken. Dit experiment werd uitgevoerd met hulp van 24 moedertaalsprekers van Georgisch. Twee afhankelijke variabelen werden gebruikt om de resultaten van het experiment te evalueren, namelijk detectiegraad (detectiepercentage) en detectie latentie (reactietijd in ms.). Het experiment ondersteunt de aanname dat harmonische groepen complexe segmenten zijn.

Reduplicatie in het algemeen laat zien wat de ongemarkeerde structuren van een taal zijn, en reduplicatie in het Georgisch (hoofdstuk 5) ondersteunt de aannames gemaakt in hoofdstuk 3: (i) het minimale woord in het Georgisch is bisyllabisch; (ii) harmonische groepen kunnen geanalyseerd worden als complexe segmenten; (iii) combinaties van C + /v/ kunnen geanalyseerd worden als complexe segmenten; (iv) clusters van obstruent + sonorant zijn de minst gemarkeerde clusters in het Georgisch; en (v) het Georgisch staat clusters van maximaal twee consonanten toe aan het begin van de stam. In het algemeen zijn consonantsequenties afgeleid.

De laatste twee hoofdstukken betreffen analyses van consonantsequenties in het Georgisch. De Gradual Consonant Analysis (GCA) wordt voorgesteld in hoofdstuk 6. De analyse verenigt paradigmatische, syntagmatische, fonetische, historische en comparatieve vormen van evidentie. De analyse van consonantale stammen in het Georgisch en het toepassen van de GCA blijken een betrouwbare basis voor het onderzoek van ‘complexe’ structuren in een taal. In de meeste gevallen blijken de lange consonantsequenties het resultaat te zijn van een zeer productieve regel van vocaaldeletie en van de vorming van complexe segmenten. Het Georgisch is een inflectioneel-agglutinatieve taal en consonantsequenties zijn in het algemeen afgeleid van CVC stammen waaraan affixen met een initiële vocaal zijn aangehecht. In die zin is de fonologische ‘complexiteit’ in de eerste plaats een gevolg van morfologische complexiteit. De structuren van het type CCC in het Georgisch lijken het resultaat van veranderingen in CVCVCV structuren, zoals gesuggereerd werd in hoofdstuk 1 en 2.

In hoofdstuk 7 beschouw ik eerdere analyses van Georgische consonantsequenties: de Headless Syllable Analysis (HSA), de Syllabified Consonant Analysis (SCA) en een analyse in het kader van Government Phonology (Toft 1999). Er worden verschillende criteria opgesteld om de GCA te vergelijken met deze eerdere analyses: (i) eenvoud of economie van beschrijving; (ii) algemeenheid of abstractheid; en (iii) adequaatheid (empirisch en verklarend). De vergelijking laat zien dat de GCA een beter begrip van de consonant sequenties biedt dan de voorgaande analyses.

De verschillen tussen de eerdere analyses en de GCA liggen mede in de concepten waarmee zij werken. In de HSA, de SCA en in Tofts analyse wordt de vormkant van taal, de fonotaxis, geïnterpreteerd in termen van formele eenheden, zoals de syllabe, de voet, de onset en de nucleus. In de GCA wordt de fonotaxis gezien als

een samenspel van drie dimensies: betekenis, vorm en structuur. Een dergelijke vernieuwende benadering van de studie van fonotactische restricties verbindt de fonologie met de andere modules van het taalsysteem. Dit levert een beter inzicht op van verschijnselen die eerder alleen als fonologisch werden beschouwd. Zeker wanneer de morfologie in beschouwing wordt genomen, wint de studie van fonotaxis aan verklarende kracht.

Een laatste verschil tussen de eerdere analyses en de GCA ligt in het feit dat de eerdere benaderingen fonotaxis als een statische, structurele component beschouwen, terwijl de GCA, onder andere door ook comparatieve en historische evidentie te gebruiken, fonotaxis als een dynamische component van taalorganisatie beschouwt.

Samenvattend worden de volgende stellingen verdedigd in deze dissertatie:

- Fonotaxis is dynamisch.
- Fonotaxis is gevoelig voor de morfologie.
- De stam is een fonologisch domein.
- Consonantale complexiteit in het Modern Georgisch is het gevolg van morfologische complexiteit, en het resultaat van processen van vocaaldeletie en van het vormen van complexe segmenten.

რეზიუმე **(Summary in Georgian)**

ამ დისერტაციის მთავარ მიზანს წარმოადგენს თანხმოვანთკომპლექსების (მაგალითად, CCC ტიპის სტრუქტურათა) ფენომენის შესწავლა. ფარდობითობისა და დამატებითობის პრინციპების გამოყენებით ჩამოყალიბებულია ჰიპოთეზა CCC და CVCVCV ტიპის სტრუქტურათა ურთიერთკავშირის შესახებ. თანხმოვანთა კომპლექსები განხილულია როგორც CVC ტიპის სტრუქტურებიდან ნაწარმოები (მეორადი) ოდენობები, სადაც CVC სიტყვის ლექსიკურ ნაწილს ანუ ფუძეს წარმოადგენს.

რაც შეეხება საკითხის ფორმალიზაციას, თანხმოვანთა ფონოტაქტიკის ადეკვატური აღწერისათვის შემოთავაზებულია ალტერნატიული პროსოდიული იერარქია, სადაც ფუძეს სიტყვასა და სეგმენტს შორის უჭირავს ადგილი. ენის სამი ძირითადი კომპონენტის: ფორმის, მნიშვნელობისა და სტრუქტურის კავშირის გათვალისწინებით, სიტყვის ფონოტაქტიკა შესწავლილია ერთდროულად სამი განზომილებიდან - (1) სეგმენტთა ხაზოვანი მიმდევრობები სიტყვის ფორმალურ ასპექტს შეადგენს; (2) მორფოლოგიური შედგენილობა სიტყვის მნიშვნელობის ასპექტს, ხოლო (3) ფონოლოგიური პრინციპები სიტყვის სტრუქტურულ ასპექტს. დისერტაციის ძირითადი ნაწილი ეძღვნება პირველ და მეორე თავებში ჩამოყალიბებული ჰიპოთეზების შემოწმებას. თანხმოვანთკომპლექსები, მაგალითად CCC ტიპის ფორმები, შესწავლილია ქართული ენის მასალაზე.

მათთვის, ვინც თვლის რომ ენა სისტემაა, ქართული ენის თანხმოვანთკომპლექსები, მაგალითად, ისეთები როგორიცაა /ფრცქვნ/, /მწვრთნ/, /ბრტყ/ სიტყვებში /ფრცქვნა/, /მწვრთნელი/ და /ბრტყელი/, წარმოადგენს ზედმეტად გრძელ და, ერთი შეხედვით, არაორგანიზებულ ოდენობებს. თანხმოვანთკომპლექსთა გასაანალიზებლად შესწავლილია თანხმოვანთა ქცევა იზოლირებულად და მიმდევრობებში. საბოლოო დასკვნების ჩამოყალიბებისას მხედველობაში მიღებულია როგორც სინქრონიული, ასევე დიაქრონიული მონაცემები. გათვალისწინებულია სხვა ქართველური ენების მასალაც.

მესამე თავში განხილულია სამი ტიპის თანხმოვანთმიმდევრობა, რომელიც შეიძლება გაანალიზდეს როგორც კომპლექსური სეგმენტი. ეს თანხმოვანთმიმდევრობებია: (1) ჰარმონიული კომპლექსები; (2) ჩქამიერი თანხმოვანი + /ვ/, და (3) /ს/ + ჩქამიერი თანხმოვანი. ამავე თავში აღნიშნულია რომ მინიმალური სიტყვა ქართულში ორმაცვლიანია და თანხმოვანთა კომპლექსების განაწილება შემოსაზღვრულია ფუძით (ლექსიკური მორფემით). მეორე თავში ჩამოყალიბებული ჰიპოთეზის მართებულობა თანხმოვანთკომპლექსთა მეორადი ბუნების შესახებ

შემოწმებული და დადასტურებულია მესამე თავში. ამავე დროს, შედარებულია თანხმოვანთა დისტრიბუციული თავისებურებანი მიმდევრობებში და დისტანციაზე. თანხმოვანთა დისტრიბუციის ამ ორ კონტექსტს აღმოაჩნდა ბევრი საერთო ნიშან-თვისება, რომელმაც დაამტკიცა იმ დებულების სისწორე, რომელიც ჩამოყალიბებული იყო პირველ თავში: "თანხმოვანთკომპლექსები მეორადნი (ნაწარმოებნი) არიან". თანხმოვანთკომპლექსები თანამედროვე ქართულში წარმოიშვა ფუძისეული ხმოვნის ამოღების შედეგად. ფუძისეული ხმოვნის დაკარგვა კი ძველსა და ახალ ქართულში ფართოდ გავრცელებული პროცესია.

მესამე თავში ჩამოყალიბებული დებულებების შესამოწმებლად მოვიშველიეთ დამატებითი მტკიცებულებანი. ფონოლოგიური არგუმენტი ჰარმონიულ ჯგუფთა კომპლექსურ სეგმენტებად განხილვის შესახებ შემოწმდა ფონეტიკური ექსპერიმენტით, პერცეფციული ტექნიკის გამოყენებით. ექსპერიმენტში მონაწილეობა მიიღო 24 პირმა, რომელთათვისაც ქართული მშობლიური ენა იყო. ექსპერიმენტის მონაცემები გაანალიზებულია ორი მაჩვენებლის გამოყენებით: რეაქციის დრო და რეაქციის სისწორე. მეოთხე თავში დეტალურადაა აღწერილი ამ ექსპერიმენტის პროცედურა და შედეგები. პერცეფციული ექსპერიმენტის მონაცემებმა მხარი დაუჭირა ფონოლოგიურ თვალსაზრისს ჰარმონიულ ჯგუფთა კომპლექსურ სეგმენტებად განხილვის შესახებ.

რედუპლიკაცია (გაორმაგება), ძირითადად, ენის არამარკირებულ სტრუქტურებს ავლენს. ქართული ენის რედუპლიკაციის მასალა მხარს უჭერს მესამე თავში ჩამოყალიბებულ დებულებებს: (1) ქართულში მინიმალური სიტყვა ორმარცვლიანია; (2) ჰარმონიული ჯგუფები შეიძლება გაანალიზდეს როგორც კომპლექსური სეგმენტები; (3) მიმდევრობა ჩქამიერი თანხმოვანი + /ვ/ შეიძლება გაანალიზდეს როგორც კომპლექსური სეგმენტი; (4) მიმდევრობა ჩქამიერი + სონანტი არის ქართული ენის ყველაზე არამარკირებული თანხმოვანთმიმდევრობა, და (5) ქართულში ფუძის დასაწყისში თანხმოვანთა მაქსიმალური რაოდენობა არის ორი. თანხმოვანთმიმდევრობანი მეორად ფენომენს წარმოადგენენ.

დისერტაციის ბოლო ორი თავი ეძღვნება ქართული თანხმოვანთკომპლექსების ანალიზს. მეექვსე თავში შემოთავაზებულია თანხმოვანთა თანმიმდევრული ანალიზი (The Gradual Consonant Analysis), რომელიც ეყრდნობა სხვადასხვა მონაცემს: პარადიგმატულს, სინტაგმატურს, ისტორიულს, ფონეტიკურსა და შედარებითს. ქართულ ოდენთანხმოვნიან ფუძეთა განხილვა და თანხმოვანთა თანმიმდევრული ანალიზი მიანიშნებს, რომ ენის შინაგანი და გარეგანი ფაქტორების გათვალისწინება წარმოადგენს ენის კომლექსურ სტრუქტურათა ბუნების შესწავლის წინაპირობას. აღმოჩნდა, რომ შემთხვევათა უმრავლესობაში ქართულის გრძელი თანხმოვანთმიმდევრობები ხმოვნის დაკარგვისა და კომპლექსურ სეგმენტთა ჩამოყალიბების პროცესების შედეგადაა მიღებული. ქართული აგლუტინაციურ-ფლექსიური ენაა და თანხმოვანთკომპლექსები ძირითადად მიიღება ფუძეზე აფიქსის დართვისას ფუძისეული ხმოვნის დაკარგის შედეგად. ამგვარად, ფონოლოგიური კომპლექსურობა განპირობებულია

მორფოლოგიური კომპლექსურობით და CCC ტიპის სტრუქტურები თანამედროვე ქართულში წარმოადგენს CVCVCV ტიპის სტრუქტურათა ტრანსფორმაციის შედეგს, როგორც ეს არგუმენტირებულია პირველსა და მეორე თავებში.

მეშვიდე თავში მიმოხილულია ბოლო დროს შემოთავაზებული სამი ტიპის ანალიზი ქართული თანხმოვანთმიმდევრობისა: ნეფვიუ (1994), ბუში (1997) და თოფტი (1999). ეს ანალიზები შეუდარდა თანხმოვანთა თანმიმდევრულ ანალიზს. შედარებისას გამოყენებულ იქნა სამი ტიპის კრიტერიუმი: (1) სიმარტივე ან ეკონომიურობა; (2) ზოგადი ხასიათი ან აბსტრაქტულობა და (3) ადეკვატურობა (ემპირიული და განმარტებითი).

განსხვავება თანხმოვანთა თანმიმდევრულ ანალიზსა და ამ სამი ტიპის ანალიზს შორის მდგომარეობს იმ ცნებებში, რომლებითაც ისინი ოპერირებენ. ნეფვიუს (1994), ბუშის (1997) და თოფტის (1999) ანალიზში ენის ფონოტაქტიკა მხოლოდ ფორმალურ ტერმინებშია ინტერპრეტირებული. მაგალითად, ისეთებში, როგორიცაა: მარცვალი (syllable), ტერფი (foot), მარცვლის დასაწყისი (onset) და ბირთვი (nucleus). თანხმოვანთა თანმიმდევრული ანალიზი კი სიტყვის ფონოტაქტიკას განიხილავს სამი ასპექტის გათვალისწინებით. ეს ასპექტებია: ფორმა, მნიშვნელობა და სტრუქტურა. თანხმოვანთა თანმიმდევრული ანალიზი ფონოლოგიას აკავშირებს ენის სხვა დონეებთან, მაგალითად, მორფოლოგიასთან. ამგვარი მიდგომა კი თანხმოვანთმიმდევრობათა ფენომენში გარკვევის უკეთეს საშუალებას იძლევა.

გარდა ამისა, განსხვავება ამ ორი ტიპის ანალიზს შორის მდგომარეობს ფონოტაქტიკისადმი ახლებურ მიდგომაში. ადრე შემოთავაზებული ანალიზებისაგან განსხვავებით, სადაც ფონოტაქტიკა განიხილებოდა როგორც ენის სტატიკური კომპონენტი, ჩვენ ვთავაზობთ ახალ მეთოდოლოგიას - სინქრონიული, დიაქრონიული და შედარებითი მეთოდების შეჯერებით ფონოტაქტიკის ენის დინამიურ კომპონენტად განხილვას.

ამგვარად, დისერტაციაში ჩამოყალიბებულია შემდეგი თეზისები:

- ფონოტაქტიკა წარმოადგენს ენის დინამიურ კომპონენტს.
- ფონოტაქტიკა მჭიდრო კავშირშია მორფოლოგიასთან.
- ფუძე ფონოლოგიური ერთეულია.
- თანმედროვე ქართულის გრძელი თანხმოვანთმიმდევრობები სიტყვის მორფოლოგიურ სტრუქტურასთანაა დაკავშირებლი. ამასთანავე, ისინი ხმოვნის დაკარგვისა და კომპლექსურ სეგმენტთა ჩამოყალიბების პროცესების შედეგადაც არის მიღებული.

Curriculum Vitae

Marika Butskhrikidze was born on 26 March 1971, in Tbilisi, Georgia. She left the 130th G. Akhvlediani secondary school in June 1988. In September 1988 she was admitted to the Department of Structural and Applied Linguistics of the Faculty of Applied Mathematics at I. Javakhishvili Tbilisi State University, from which she graduated with honours in June 1993 with a degree in Linguistics. During the same period, 1998–1993, she was a guest student at the Department of Oriental Studies (Persian language). From 1993 to 1996 she worked as a senior assistant at the Laboratory of Kartvelian Languages, Sulkhan-Saba Pedagogical Institute. During the same period she also worked as a research fellow at the Department of Language Typology, G. Tsereteli Institute of Oriental Studies, Georgian Academy of Sciences. As a result of this work, in January 1998 she defended her dissertation on the "Typology of fixed accent" at the Tbilisi State University, and graduated as Candidate of Philological Sciences. During the period 1994–1996 she studied at the Faculty of Higher Courses of Simultaneous Translation, I. Chavchavadze Institute of Foreign Languages. From September 1996 to August 1997 she had a graduate position in the Theoretical Linguistics Doctoral Programme at the Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest. From 1 September 1997 until 31 January 2002 she worked on a project "The consonant phonotactics of Georgian" at the University of Leiden Centre for Linguistics. This dissertation is the result of research carried out during this period.